AWARD NUMBER: W81XWH-13-1-0319

TITLE: A Novel Approach to Assay DNA Methylation in Prostate Cancer

PRINCIPAL INVESTIGATOR: Jindan Yu

CONTRACTING ORGANIZATION: Northwestern University
303 E. Superior St.
Chicago, IL 60611

REPORT DATE: December 2017

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. REPORT DATE | 2. REPORT TYPE | 3. DATES COVERED |
|---|---|---|
| December 2017 | Final | 30Sept2013 - 29Sept2017 |

**4. TITLE AND SUBTITLE**

A Novel Approach to Assay DNA Methylation in Prostate Cancer

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**
W81XWH-13-1-0319

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

Jindan Yu

E-Mail: jindan-yu@northwestern.edu

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Northwestern University
303 E. Superior St.
Chicago, IL 60611

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION / AVAILABILITY STATEMENT**

Approved for Public Release; Distribution Unlimited

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

Previous studies of DNA methylation at 5-position of cytosine (5mC) have led to the discovery of useful methylation biomarkers for prostate cancer diagnosis and prognosis, some of which are being developed into clinical tests. However, several seminal studies have recently reported that DNA methylation (5mC) can be de-methylated by the TET proteins resulting in 5-hydroxylmethylation (5hmC), which plays functional roles distinct from 5mC but yet is indistinguishable from 5mC by a majority of existing methylation assays. Developing enabling assays that measure 5mC and 5hmC specifically might significantly improve the performance of methylation biomarkers. We have shown that MeDIP-seq and 5hmC-seq as optimal approaches to globally detect 5mC and 5hmC, respectively. We have then performed genome-wide mapping of 5mC and 5hmC in 33 samples, including 11 primary benign prostate tissues, 11 localized prostate cancer tissues, and 6 castration-resistant prostate cancer tissues. Bioinformatics analysis revealed a molecular crosstalk among TET1, FOXA1-mediated enhancers, and AR signaling and identified cancer-specific methylation biomarkers which will be further validated in larger sample sets.

**15. SUBJECT TERMS**

DNA methylation 5mC, DNA hydroxymethylation 5hmC, biomarker, 5hmCSL-PCR, 5mCSL-PCR, BS-seq, TAB-seq, cancer-specific methylation

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | USAMRMC |
| Unclassified | Unclassified | Unclassified | Unclassified | 65 | 19b. TELEPHONE NUMBER *(include area code)* |

## Table of Contents

**Page**

1. **INTRODUCTION:**

Previous studies of DNA methylation at 5-position of cytosine (5mC) have led to the discovery of useful methylation biomarkers for prostate cancer diagnosis and prognosis, some of which are being developed into clinical tests. However, several seminal studies have recently reported that DNA methylation 5mC can be de-methylated by the TET proteins resulting in 5-hydroxylmethylation (5hmC), which plays distinct functional roles of 5mC but yet is indistinguishable from 5mC by a majority of current assays. Developing enabling assays that measure 5mC and 5hmC specifically might significantly improve the performance of methylation biomarkers. Our goal is thus to develop and apply 5mC- or 5hmC-specific assays to study prostate cancer (PCa) methylome. Our hypothesis is that 5hmC and 5mC play distinct and important roles in PCa and cancer-specific 5hmC and 5hmC signatures may be useful PCa biomarkers. To test this hypothesis, we propose two Specific Aims:

(1) Develop 5mC- and 5hmC-specific assays for DNA methylation analysis in PCa.

(2) Obtain genome-wide maps of 5mC and 5hmC distribution in prostate cancer.

2. **KEYWORDS:** DNA methylation 5mC, DNA hydroxymethylation 5hmC, biomarker, 5hmCSL-PCR, 5mCSL-PCR, BS-seq, TAB-seq, cancer-specific methylation

3. **ACCOMPLISHMENTS:**

- **What were the major goals of the project?** The Table below listed the goals/target dates of the project as in the approved SOW. A new column was added on the right showing actual completion dates and the percentage of completion.

| **Specific Aim 1. To develop 5mC- and 5hmC-specific assays for DNA methylation analysis in PCa** | **Timeline** | **Site 1** | **Site 2** | **% of completion** | **Actual complete dates** |
|---|---|---|---|---|---|
| **Major Task 1 : Develop 5hmCSL-PCR assay** | Months | | | | |
| Subtask 1: optimize 5hmCSL-PCR in a panel of prostate cell lines using previously known methylated regions | 1-6 | Dr. Yu | Dr. He | 100% | 9/1/2015 |
| Subtask 2: Compare 5hmCSL results with hMeDIP. | 6-12 | Dr. Yu | | 100% | 12/31/2015 |
| Subtask 3: Use TAB-pyro to validate 5hmCSL results. | 6-12 | Dr. Yu | | 50% | 9/30/2017 |
| Local IRB non-human subject Approval | 1-3 | | | 100% | 3/22/2014 |
| Milestone Achieved: HRPO Approval | 6 | | | 100% | 3/22/2014 |
| **Major Task 2: Develop 5mCSL-PCR assay** | | | | | |
| Subtask 1: optimize 5mCSL-PCR in a panel of prostate cell lines using previously known | 3-9 | Dr. Yu | Dr. He | 100% | 2/1/2016 |

| | | | | | |
|---|---|---|---|---|---|
| methylated regions | | | | | |
| Subtask 2: Compare 5mCSL results with MeDIP. | 6-12 | Dr. Yu | | 100% | 6/31/2014 |
| Subtask 3: Use BS-pyro and TAB-pyro to validate 5mCSL results. | 9-18 | Dr. Yu | | 100% | 9/30/2017 |
| **Major Task 3: Assess methylation biomarkers in prostate cancer specimens.** | | | | | |
| Subtask 1: analyze 5mC and 5hmC level of selected genes using DNA from 90 prostate cancer tissues. | 12-18 | Dr. Yu | | 60% | 9/30/2017 |
| **Specific Aim 2: Obtain genome-wide map of 5mC and 5hmC distribution in prostate cancer** | | | | | |
| **Major Task 4: Genome-wide 5(h)mCSL-Seq** | | | | | |
| Subtask 1: perform 5(h)mCSL-Seq in 3 cell lines. | 9-18 | Dr. Yu | Dr. He | 100% | 9/30/2017 |
| Subtask 2: identify cancer-specific 5(h)mC loci. | 12-30 | Dr. Yu | | 80% | 9/30/2017 |
| **Major Task 5: BS-Seq and TAB-Seq** | | | | | |
| Subtask 1: BS-Seq and TAB-Seq in 3 cell lines | 24-30 | Dr. Yu | Dr. He | 100% | 9/30/2017 |
| Subtask 2: identify 5(h)mC-rich loci and cross-validation using different assays. | 30-36 | Dr. Yu | | 60% | 9/30/2017 |
| Subtask 3: analyze cancer-specific 5(h)mC biomarker using DNA from 90 prostate cancer tissues. | 30-36 | Dr. Yu | | 60% | 9/30/2017 |

- **What was accomplished under these goals?**

1) Major activities in this reporting period include:

1. Additional sequencing was performed on 5hmCSL-seq human prostate samples previously sequenced for more robust and accurate data for bioinformatic analysis.
2. Bioinformatic analysis of human prostate sample methylomes and hydroxymethylomes to identify prostate cancer progression associated epigenetic changes.
3. Identification of prostate cancer tissues presenting recurrence and non-recurrence of disease for progression associated biomarker development.
4. Develop collaborations for 5hmCSL-seq detection in cell-free DNA from serum for CRPC monitoring and treatment responses.
5. Determine how TET1 impacts AR signaling.

2) Specific objectives: (A) Optimize existing and novel methylation assays; (B) obtain genome-wide maps of 5mC and 5hmC.

3) Significant results or key outcomes:

**Specific Aim 1. To develop 5mC- and 5hmC-specific assays for DNA methylation analysis in PCa**

**Major Task 1: Develop 5hmCSL-PCR assay**

**Subtask 1: optimize 5hmCSL-PCR in a panel of prostate cell lines using previously known methylated regions**

Methods: Genomic DNA was isolated from a panel of prostate cell lines including LNCaP, 22Rv1, RWPE, VCaP, PrEC, DU145, and BPH1. For 5hmCSL, an azide-containing modified glucose was

added onto 5hmC molecules utilizing the T4 bacteriophage β-gycosyltransferase (β-GT). By click chemistry, a biotin tag was then installed on the azide group, thus labeling 5hmC, which was subsequently pulled down as mediated by streptavidin. Once the 5hmC-containing DNA were successfully pulled down, we carried out QPCR analysis of several regions that have previously been shown to be methylated in prostate cancer.

Results: QRT-PCR results (**Fig.1, left two columns of each gene**) showed that 5hmC levels are in generally lower than 5mC for most of the genes (except RASSF) in most cell lines. By contrast, GAPDH has sometimes higher 5hmC than the tested genes, being consistent with its abundant expression.

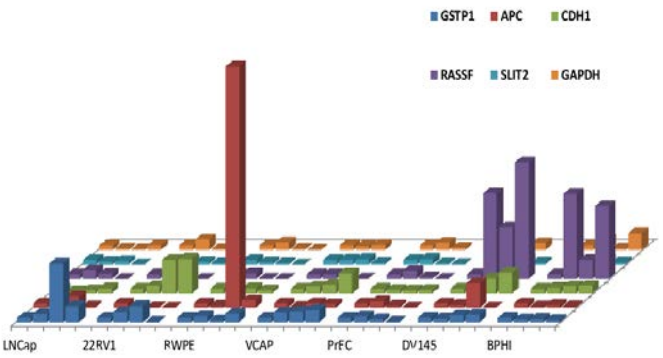



**Fig.1 5mCSL and 5hmCSL pull down using DNA from various prostate cancer cell lines.** The enriched DNA was then subjected to QPCR analysis of 5 previously reported methylated genes in prostate cancer (GSTP1, APC, CDH1, RASSF2, SLIT2) and a control gene (GAPDH). For each gene at each condition, shown on the left are **5hmC** and on the right are **5mC** with duplicates.

Conclusion: 5hmC can be detected using 5hmCSL-PCR.

**Subtask 2: Compare 5hmCSL results with hMeDIP.**

To compare the performance of 5hmCSL and hMeDIP, we conducted 5hmCSL-qPCR and hMeDIP-qPCR on a set of genes and found that the former is much more sensitive and accurate in capturing 5hmC (**Fig.2**). The 5hmCSL-qPCR assay, as expected, showed reduced 5hmC levels at target methylated regions following knockdown of TET1, the DNA demethylase that converts 5mC to 5hmC. This is consistent with the fact that 5hmC only exists on a small percentage (~2.5%) of Cs and may be insufficient to enrich by antibody-based approach (the hMeDIP assay).

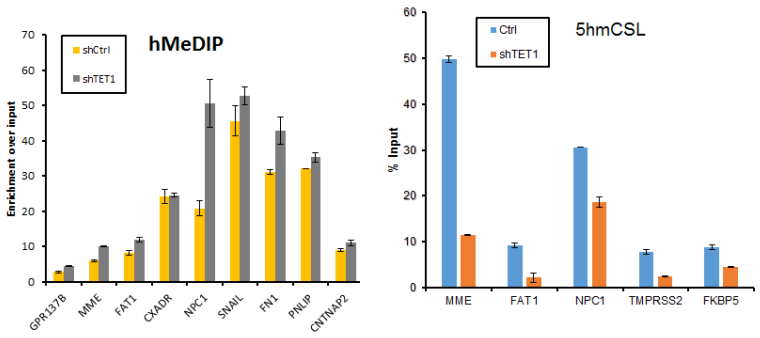



**Fig. 2**: 5hmCSL-PCR is more accurate in determining 5hmC level than hMeDIP. LNCaP cells with control and TET1 knockdown were subjected to hMeDIP and 5hmCSl followed by qPCR analysis.

**Subtask 3: Use TAB-pyro to validate 5hmCSL results.**

We have performed BS-seq and TAB-seq (genome-wide assays of BS-pyro and TAB-pyro) in LNCaP cells. To validate 5hmCSL results and demonstrate the accuracy of 5hmCSL sequencing analysis of 5hmC sites, we will compare the TAB-seq data with 5hmCSL-seq data (**Table 1**). Bioinformatics analysis of the data is yet to be completed.

**Table 1**: Validation of 5hmCSL-seq assay using BS-seq and TAB-seq

| Cell line | Condition | Method | barcode | # seq reads |
|---|---|---|---|---|
| LNCaP | control | 5hmC-seq | 701-503 | 7,591,719 |
| LNCaP | DMSO | 5hmC-seq | 703-505 | 5,365,282 |
| LNCaP | shGIPZ TAB | TAB-seq | TATAAT | 306,755,767 |
| LNCaP | shFOXA1 TAB | TAB-seq | TCGAAG | 273,079,194 |

**Major Task 2: Develop 5mCSL-PCR assay**

**Subtask 1: optimize 5mCSL-PCR in a panel of prostate cell lines**

**using previously known methylated regions**

Methods: Genomic DNA was isolated from a panel of prostate cell lines including LNCaP, 22Rv1, RWPE, VCaP, PrEC, DU145, and BPH1. For 5mCSL, 5hmC was first protected through glucosylation mediated by the T4 bacteriophage β-gycosyltransferase (β-GT). Then using the mouse Tet1 catalytic domain, 5mC was converted into 5hmC and simultaneously glucosylated with a modified glucose moiety (6-N3-glucose), which allows it to be selectively labelled with biotin in a subsequent Huisgen cycloaddition step (click chemistry). Lastly, the biotin-tagged 5mC fragments were pulled down. Once the 5mC-containing DNA were successfully pulled down, we carried out QPCR analysis of several regions that have previously been shown to be methylated in prostate cancer.

Results: QRT-PCR results of 5mC (**Fig.1, right two columns of each gene**) showed varying degree of 5mC methylation of the positive control genes in cancer cell lines. GSTP1 was shown to be methylated mainly in LNCaP, 22RV1, and VCaP cells. APC is methylated in LNCaP and DU145. CDH1 appear to be methylated in a number of prostate cancer cell line including DU145, VCaP, and 22Rv1. RASSF2 is methylated strongly in DU145 and BPH1. The methylation pattern of SLIT2 was not very clear based on this experiment. There could be a problem with the SLIT2 primers in this experiment. By contrast, the negative-control gene GAPDH is barely methylated in any cell lines tested.

Conclusion: We observe an overall trend of success of 5mCSL-PCR and 5hmCSL-PCR in detecting 5mC and 5hmC methylation, respectively, in prostate cancer.

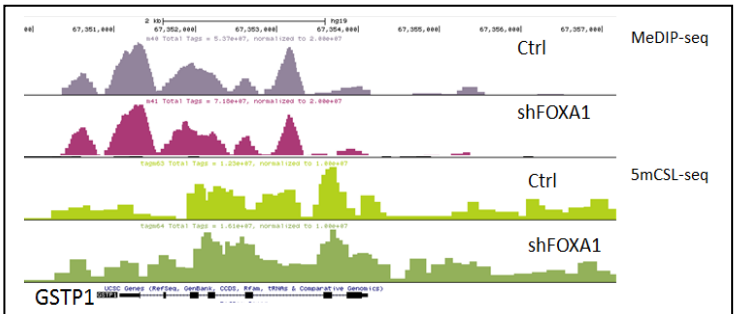



**Fig.3**: MeDIP-seq outperformed 5mCSL-seq. LNCaP cells with control and FOXA1 knockdown were subjected to MeDIP and 5mCSL pull down followed by NGS analysis. The enriched peaks around the GSTP1 gene are shown.

**Subtask 2: Compare 5mCSL results with MeDIP.** In order to compare the effectiveness of 5mCSL assay in relative to MeDIP, we performed next-generation sequencing of the enriched DNA. We found that with comparable amount of total sequencing reads, 5mCSL-seq leads to fewer and indistinct peaks when compared to MeDIP-seq which render sharp and clear peaks of enrichment (**Fig.3**). This may be due to the fact that 5mC is quite abundant and the level of modified 5mC is sufficient for antibody-based enrichment. On the other hand, 5mCSL requires blockade of original 5hmC and conversion of 5mC to new 5hmC for subsequent chemical-linked pull down, a process that may result in substantial loss of the material.

**Table 2**: Validation of MeDIP-seq assay using BS-seq and TAB-seq

| Cell line | Condition | Method | barcode | # seq reads |
|---|---|---|---|---|
| LNCaP | shGIPZ BS | BS-seq | GGTAGC | 235,589,814 |
| LNCaP | shGIPZ TAB | TAB-seq | TATAAT | 306,755,767 |
| LNCaP | control | MeDIP-seq | ACAGTG | 15,886,470 |

**Subtask 3: Use BS-pyro and TAB-pyro to validate 5mCSL results.**

As MeDIP-seq out-performed 5mCSL-seq, it is no longer needed to compare 5mCSL based assay with pyrosequencing. Instead, we will evaluate MeDIP-seq results through comparative analysis with BS-seq and TAB-seq data (**Table 2**). We expect methylated regions detected by MeDIP-seq to contain signals in BS-seq but not TAB-seq.

**Major Task 3: Assess methylation biomarkers in prostate cancer specimens.**

**Subtask 1: analyze 5mC and 5hmC level of selected genes using DNA from 90 PCatissues.**

We initially proposed to analyze a set of known methylated genes/loci in 90 benign adjacent and prostate cancer tissues. However, as our data (**Fig.1**) did not support their being promising methylation

biomarkers, we decided to obtain a cancer-specific 5mC or 5hmC signature through global 5mC and 5hmC profiling in a large set of primary prostate cancer specimens as detailed below in Aim 2.

**Conclusion of Specific Aim 1**: we have optimized out assays and we will use MeDIP-seq for genome-wide analysis of 5mC and 5hmCSL-seq for genome-wide mapping of 5hmC.

**Table 3**: 5mCSL-seq and 5hmCSL-seq in PCa cell lines.

| Cell line | Condition | Method | Barcode | Name | #seq reads |
|---|---|---|---|---|---|
| PrEC LHS | Normal | 5mC pull down | CGATGT | A1 | 27,462,926 |
| LNCaP | Normal | 5mC pull down | TGACCA | A2 | 28,687,039 |
| PC-3M | Normal | 5mC pull down | ACAGTG | A3 | 24,459,274 |
| PrEC LHS | Normal | 5hmC pull down | GCCAAT | A4 | 29,953,414 |
| LNCaP | Normal | 5hmC pull down | CAGATC | A5 | 25,075,174 |
| PC-3M | Normal | 5hmC pull down | CTTGTA | A6 | 30,563,514 |

**Specific Aim 2: Obtain genome-wide map of 5mC and 5hmC distribution in prostate cancer**

**Major Task 4: Genome-wide 5(h)mCSL-Seq**

**Subtask 1: perform 5(h)mCSL-Seq in 3 cell lines.**

We have performed 5mCSL-seq and 5hmCSL-seq in three prostate cell lines as proposed, representing benign, androgen-dependent, and androgen-independent prostate cancers (**Table 3**). During data analysis, we found that 5mC is depleted, whereas 5hmC enriched, at enhancer regions marked by FOXA1 binding in the AR+, FOXA1+ LNCaP cells, but not in the PrEC and PC-3M cells (**Fig.4**). We thus hypothesized that FOXA1 may be a critical regulator of TET1 and/or DNA demethylation. To investigate this, we have examined how FOXA1 regulates TET1 expression and enhancer function. The key findings are summarized below and the manuscript reporting the results has been published in Nucleic Acids Res (**Appendix 1**).

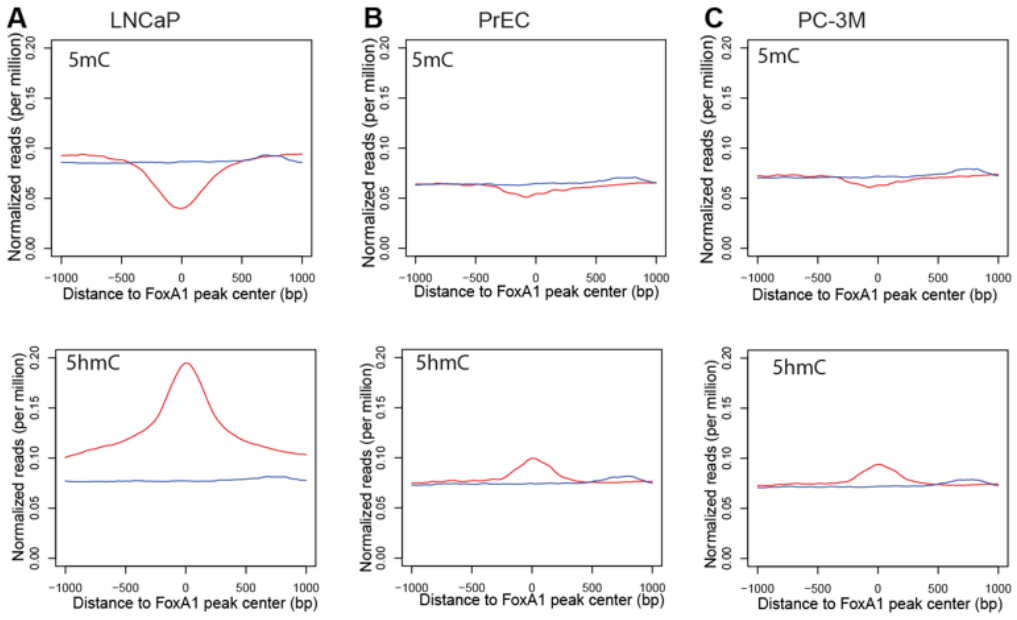



**Fig. 4**: Intensity plots showing 5mC and 5hmC enrichment around FOXA1 binding sites (±1 kb) in LNCaP (A), PrEC (B), and PC-3M cells (C). MeDIP and 5hmCSL were performed using genomic DNA extracted from LNCaP, PrEC and PC-3M cell lines. Enriched DNA was made into libraries and subjected to deep sequencing. The read intensities of were evaluated relative to FOXA1 binding sites in LNCaP.

<u>Investigate how FOXA1 regulates lineage-specific enhancers through modulating TET1.</u>

**Major findings:**

1. FOXA1 contributes to enhancer activation through epigenetic modifications (**Figure 1 of Appendix 1**).
2. FOXA1 induces TET1 gene expression (**Figure 2 of Appendix 1**).
3. TET1 is a direct transcriptional target of FOXA1 (**Figure 3 of Appendix 1**).
4. FOXA1 and TET1 proteins physically interact (**Figure 4 of Appendix 1**).
5. TET1 mediates active epigenetic modification at FOXA1-bound enhancers (**Fig. 5 of Appendix 1**).
6. TET1 is required for FOXA1 recruitment to lineage-specific enhancers (**Figure 6 of Appendix 1**).

**Conclusion (Fig.5 on the right)**: FOXA1 protein occupies at an intragenic enhancer of the TET1 gene to directly induce its expression. Through direct interaction with FOXA1 protein, TET1 modulates DNA demethylation and subsequently H3K4 methylation and H3K27 acetylation at FOXA1-target enhancers, which in turn facilitates FOXA1 recruitment. Thus, FOXA1 and TET1 form a positive feedback loop in lineage-specific enhancer activation. FOXA1 is not only a reader, but also a writer of epigenetic signatures at lineage-specific enhancers.

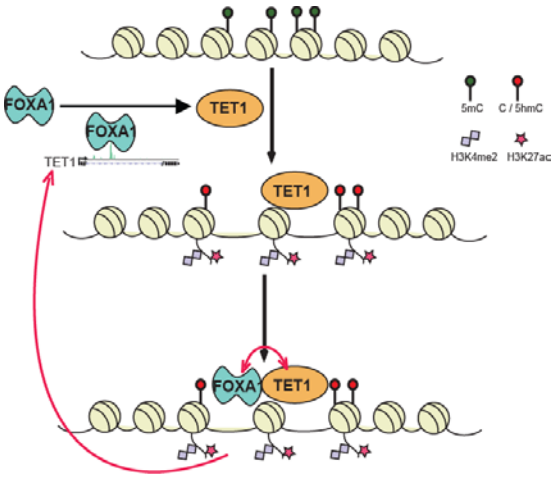



**Fig.5**. A model depicting the molecular crosstalk between FOXA1 and TET1.

**Subtask 2: identify cancer-specific 5(h)mC loci.**

2.1. In order to identify cancer-specific 5mC and 5hmC methylation. We have successfully carried out genome-wide mapping of 5mC and 5hmC in a cohort primary tissue specimens as listed below (**Table 4**). We performed additional 5hmCSL-Seq of the prepared sequencing libraries to bolster analysis and acquire sufficient coverage for downstream bioinformatics analysis. These are in addition to MeDIP-seq that was performed previously on these tissues at sufficient coverage and did not require additional sequencing reads.

**Table 4. Next-generation sequencing analysis of DNA methylation (5mC and 5hmC) in primary specimens.**

| Sample ID | Cell Line | Treatment | Antibody | Barcode | Previous Reads | Total Reads |
|---|---|---|---|---|---|---|
| 9068-21 | normal | tissue | 5hmC-seq | 701-502 | 5,464,789 | 8,167,722 |
| 31443 | normal | tissue | 5hmC-seq | 702-503 | 3,379,152 | 5,667,938 |
| 31155 | normal | tissue | 5hmC-seq | 703-504 | 4,328,516 | 7,092,330 |
| 9050-24 | normal | tissue | 5hmC-seq | 704-505 | 6,007,179 | 8,964,246 |
| 9102-21 | normal | tissue | 5hmC-seq | 705-506 | 3,308,939 | 4,846,891 |
| 30833 | normal | tissue | 5hmC-seq | 706-507 | 4,526,880 | 22,560,548 |
| 31677 | normal | tissue | 5hmC-seq | 707-508 | 3,682,219 | 24,941,572 |
| 38597 | normal | tissue | 5hmC-seq | 708-517 | 4,231,648 | 28,116,117 |
| 49246 | normal | tissue | 5hmC-seq | 709-502 | 5,304,758 | 31,396,064 |
| 38545 | normal | tissue | 5hmC-seq | 710-503 | 2,925,850 | 22,352,741 |
| Input1 | normal | tissue | 5hmC-seq | 703-503 | 2,118,178 | |
| 9068-23 | G7/91-100% | tissue | 5hmC-seq | 711-504 | 4,406,930 | 14,884,673 |
| 31154 | G9/61-70% | tissue | 5hmC-seq | 712-505 | 9,075,967 | 10,841,809 |
| 31444 | G9/71-80% | tissue | 5hmC-seq | 701-506 | 12,399,696 | 14,002,315 |
| 9050-23 | G7/51-60% | tissue | 5hmC-seq | 702-507 | 6,456,052 | 19,812,575 |
| 9102-24 | G9/20-30% | tissue | 5hmC-seq | 703-508 | 11,646,542 | 14,150,732 |
| 30831 | G9/41-50% | tissue | 5hmC-seq | 704-517 | 8,068,573 | 29,712,746 |
| 31678 | G9/61-70% | tissue | 5hmC-seq | 705-502 | 6,236,421 | 30,374,781 |
| 38600 | G9/71-80% | tissue | 5hmC-seq | 706-503 | 5,370,120 | 26,699,546 |
| 49245 | G9/61-70% | tissue | 5hmC-seq | 707-504 | 3,622,852 | 28,463,538 |
| 38547 | G9/21-30% | tissue | 5hmC-seq | 708-505 | 6,004,212 | 29,060,417 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Input 2 | cancer | tissue | 5hmC-seq | 704-504 | 3,599,877 | |
| 31-27 | CRPC | tissue | 5hmC-seq | 709-506 | 6,226,398 | 36,646,315 |
| 37-54 | CRPC | tissue | 5hmC-seq | 710-507 | 3,464,594 | 25,446,069 |
| 30-28 | CRPC | tissue | 5hmC-seq | 711-508 | 6,442,382 | 29,088,378 |
| 28-34 | CRPC | tissue | 5hmC-seq | 712-517 | 3,287,285 | 27,165,719 |
| 33-89 | CRPC | tissue | 5hmC-seq | 702-502 | 2,385,396 | 25,515,146 |
| Input3 | CRPC | tissue | 5hmC-seq | 705-505 | 3,153,956 | |
| 38597 | Normal | tissue | MeDIP-seq | GGCTAC | 25,813,531 | |
| 30-28 | CRPC | tissue | MeDIP-seq | ATCACG | 31,184,661 | |
| 33-89 | CRPC | tissue | MeDIP-seq | TTAGGC | 23,055,143 | |
| 9050-23 | G7/51-60% | tissue | MeDIP-seq | CAGATC | 44,933,849 | |
| 9102-24 | G9/20-30% | tissue | MeDIP-seq | ACTTGA | 72,867,001 | |
| 30831 | G9/41-50% | tissue | MeDIP-seq | GATCAG | 29,065,739 | |
| 27 | LNCaP | control | MeDIP-seq | ACAGTG | 15,886,470 | |
| 28 | LNCaP | shTET1 | MeDIP-seq | GCCAAT | 20,621,914 | |
| 30833 | normal | tissue | MeDIP-seq | 706-507 | 18,033,668 | |
| 31677 | normal | tissue | MeDIP-seq | 707-508 | 21,259,353 | |
| 38597 | normal | tissue | MeDIP-seq | 708-517 | 23,884,469 | |
| 49246 | normal | tissue | MeDIP-seq | 709-502 | 26,091,306 | |
| 38545 | normal | tissue | MeDIP-seq | 710-503 | 19,426,891 | |
| Input1 | normal | tissue | MeDIP-seq | 703-503 | 21,826,000 | |
| 30831 | G9/41-50% | tissue | MeDIP-seq | 704-517 | 21,644,173 | |
| 31678 | G9/61-70% | tissue | MeDIP-seq | 705-502 | 24,138,360 | |
| 38600 | G9/71-80% | tissue | MeDIP-seq | 706-503 | 21,329,426 | |
| 49245 | G9/61-70% | tissue | MeDIP-seq | 707-504 | 24,840,686 | |
| 38547 | G9/21-30% | tissue | MeDIP-seq | 708-505 | 23,056,205 | |
| 31-27 | CRPC | tissue | MeDIP-seq | 709-506 | 30,419,917 | |
| 37-54 | CRPC | tissue | MeDIP-seq | 710-507 | 21,981,475 | |
| 30-28 | CRPC | tissue | MeDIP-seq | 711-508 | 22,645,996 | |
| 28-34 | CRPC | tissue | MeDIP-seq | 712-517 | 23,878,434 | |
| 33-89 | CRPC | tissue | MeDIP-seq | 702-502 | 23,129,750 | |
| 30833 | normal tissue | tissue | MeDIP-seq | M198 | 55,178,693 | |
| 37-54 | CRPC tissue | tissue | MeDIP-seq | M201 | 39,691,324 | |
| 28-34 | CRPC tissue | tissue | MeDIP-seq | M203 | 42,248,520 | |
| 9068-23 | PC tissue | tissue | MeDIP-seq | M205 | 41,202,546 | |
| 31154 | PC tissue | tissue | MeDIP-seq | M206 | 32,865,094 | |
| 31444 | PC tissue | tissue | MeDIP-seq | M207 | 42,734,690 | |

2.2. After acquiring sufficient sequencing coverage we performed in depth bioinformatic analysis to identify cancer-specific and CRPC-specific 5hmC loci. Using a tissue group based analysis, we scanned the genome for regions of differential 5hmC loci between 1. Normal and Cancer tissues and 2. Cancer and CRPC tissues. We discovered recurrent alterations commonly found in prostate cancer tissues **(Fig.6)**.
We also found extensive 5hmC alterations in the transition from primary PCa to CRPC **(Fig.7)**. CRPC samples exhibit significant heterogeneity, the cause of which is under current investigation by our lab.

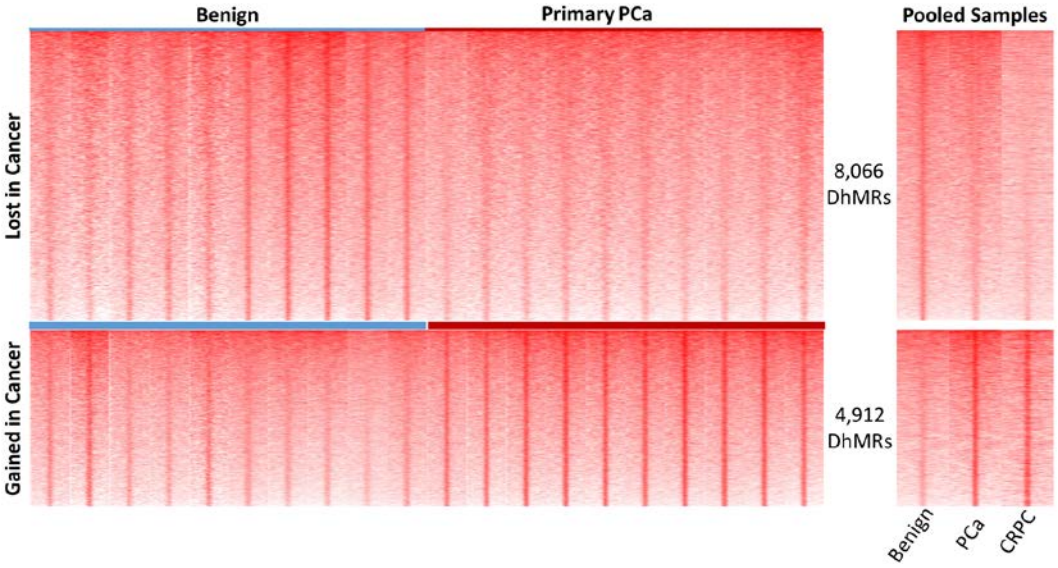



**Fig.6. Heatmaps of 5hmC alterations common among independent primary prostate cancer samples.** 5hmC levels at individual loci are depicted for each sample (left). In PCa samples, 8,066 loci exhibited loss of 5hmC (top) and 4,912 loci exhibited gains of 5hmC (bottom). The pooled sample heatmap exhibits the pooled profile of these samples (including CRPC) at these recurrently altered 5hmC regions (right)

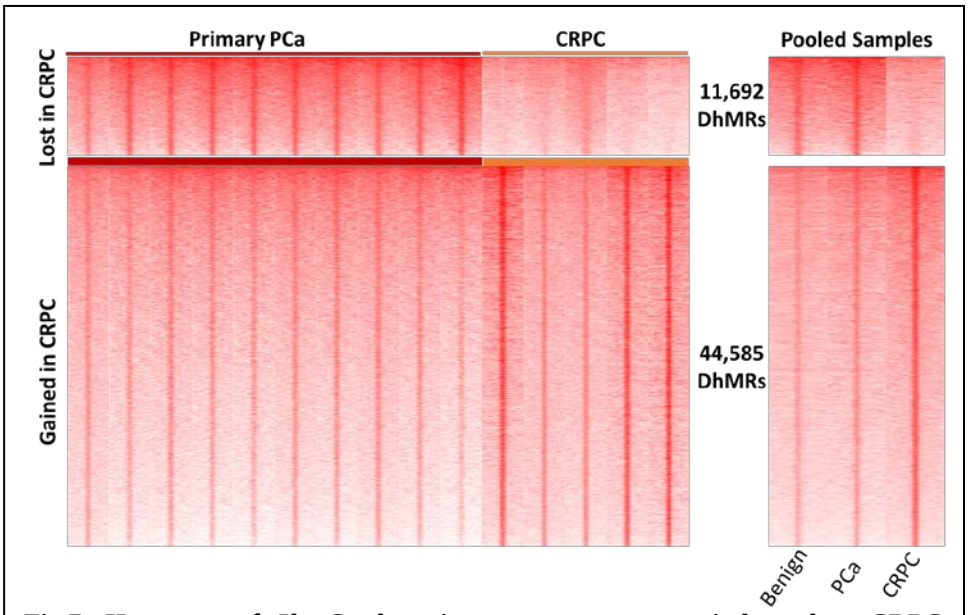



**Fig.7. Heatmaps of 5hmC alterations common among independent CRPC samples.** 5hmC levels at individual loci are depicted for each sample (left). In CRPC samples, 11,692 loci exhibited loss of 5hmC (top) and 44,585 loci exhibited gains of 5hmC (bottom). The pooled sample heatmap exhibits the pooled profile of these samples (including Benign) at these recurrently altered 5hmC regions (right)

Several downstream bioinformatic analyses have demonstrated the relevance of the detected 5hmC alterations to the oncogenic process. First, gene ontology analysis of KEGG pathways using genes containing 5hmC alterations demonstrated an enrichment for genes associated with cancer pathways for both PCa **(Table 5)** and CRPC **(Table 6)** associated 5hmC alterations. In addition, disease specific pathway analysis demonstrated an enrichment for "prostatic neoplasm" associated genes in both analyses **(not shown)**.

Second, analyzing the base pair sequence of 5hmC altered regions through motif analysis and transcription factor binding site locations yielded an enrichment of FOXA1 and AR binding sites in regions where 5hmC increased with disease (Benign to PCa and PCa to CRPC, 5hmC gain regions). This data supports our previous findings that the TET1 enzyme is recruited by FOXA1 to catalyze 5hmC at FOXA1 binding sites in prostate cells.

Third, we have found 5hmC is enriched around AR binding sites with disease progression. Using published datasets identifying cancer-specific AR binding events in patient tissues (Pomerantz/Freedman paper – citation needed? – GSE56288), we analyzed the distribution of 5hmC at human tumor-specific AR binding sites. We observed a progressive increase of 5hmC near these ARBS with disease **(Fig.8)**.

Taken together, these analyses demonstrate the relevance of 5hmC as an active participant in prostate cancer disease progression. 5hmC alterations are enriched for in genes known to participate in oncogenesis. Further, these data suggest that 5hmC is important to aberrant AR signaling, the alteration of normal AR signaling is a crucial step in prostate carcinogenesis. The manuscript describing these results and bioinformatics analysis is currently under preparation.

**Table 5: KEGG Pathway analysis of genes containing 5hmC alterations identified by comparing PCa to benign tissues**

| ID | Name | #Gene | FDR |
|---|---|---|---|
| hsa04520 | Adherens junction - Homo sapiens | 49 | 9.86E-08 |
| hsa04360 | Axon guidance - Homo sapiens | 91 | 9.43E-07 |
| hsa04144 | Endocytosis - Homo sapiens | 122 | 2.24E-06 |
| hsa05200 | Pathways in cancer - Homo sapiens | 171 | 6.08E-06 |
| hsa04510 | Focal adhesion - Homo sapiens | 97 | 1.32E-05 |

**Table 4: KEGG Pathway analysis of genes containing 5hmC alterations identified by comparing PCa to CRPC tissues**

| ID | Name | #Gene | FDR |
|---|---|---|---|
| hsa04510 | Focal adhesion - Homo sapiens | 89 | 2.3E-13 |
| hsa04360 | Axon guidance - Homo sapiens | 78 | 5.5E-12 |
| hsa04015 | Rap1 signaling pathway - Homo sapiens | 84 | 4.7E-10 |
| hsa05200 | Pathways in cancer - Homo sapiens | 126 | 1.1E-07 |
| hsa04810 | Regulation of actin cytoskeleton - Homo sapiens | 77 | 6.4E-07 |

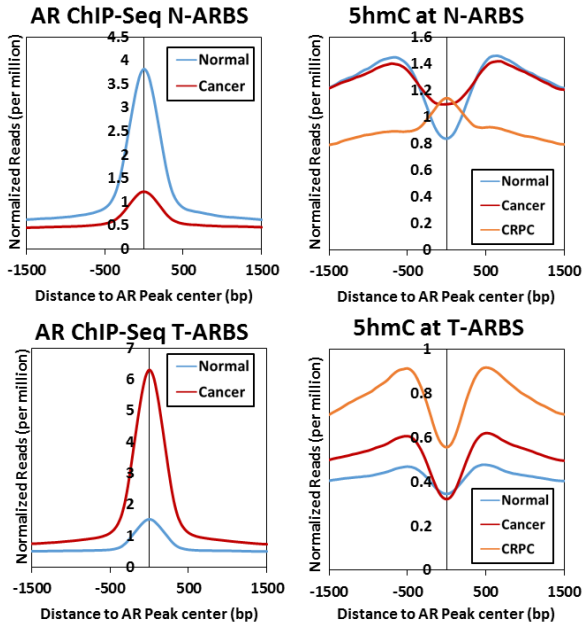



**Fig. 8. 5hmC at AR binding sites (ARBS) identified using human prostate tissues.** 5hmC density (top-right and bottom-right) at prostate tissue AR binding sites (top-left and bottom-left. At tumor-specific ARBS (T-ARBS, bottom-left), a progressive increase in 5hmC is observed with disease progression.

2.3 As we observed increased 5hmC at ARBS in our 5hmCSL-Seq analyzed human prostate tissues **(Fig.8),** we sought to examine whether TET1 regulates AR expression or signaling. **O**ur characterization of the TET1-FOXA1-5hmC pathway led to in depth analysis of the 5hmC catalyzing enzyme TET1. We identified a novel isoform of TET1 specifically expressed in AR positive cell lines. This TET1 isoform was recently confirmed as a cancer specific transcript overexpressed in cancer (Issa paper – PubMed ID: 28531272). Our preliminary data indicates TET1 is positively associated with AR expression, where knockdown of TET1 and it's novel isoform leads to a decrease in AR expression and downstream

signaling. We also observe overexpression of TET1 leads to increased binding at known ARBS, suggesting TET1 participates directly in AR downstream signaling. These studies are important in understanding aberrant AR signaling, in addition they may shed light on the mechanism of increased 5hmC at ARBS in disease progression.

**Major Task 5: BS-Seq and TAB-Seq**

**Subtask 1: BS-Seq and TAB-Seq in 3 cell lines**

We have performed this in one cell line (LNcaP) with control and knockdown of FOXA1 (which is to deregulate 5hmC level through modulating TET1) as shown below. In order to obtain base-level methylation information, we may need to do deeper sequencing to get 1 billion reads per sample.

| Cell line | Condition | Method | barcode | Sample | # seq reads |
|---|---|---|---|---|---|
| LNCaP | shGIPZ BS | BS-seq | GGTAGC | MK126 | 235,589,814 |
| LNCaP | shFOXA1 BS | BS-seq | ATGAGC | MK127 | 288,264,046 |
| LNCaP | shGIPZ TAB | TAB-seq | TATAAT | MK128 | 306,755,767 |
| LNCaP | shFOXA1 TAB | TAB-seq | TCGAAG | MK129 | 273,079,194 |

**Subtask 2: identify 5(h)mC-rich loci and cross-validation using different assays.**

During the grant support period, we write a review article on the regulation of enhancers and disease genes in prostate cancer (**Appendix 2**). Further, we have identified an enhancer-regulated (through FoxA1 and potentially TET1) lncRNA HOTAIR regulates AR signaling in prostate cancer and estrogen receptor signaling in breast cancer (**Appendix 3-4**). This grant support was thus acknowledged in the published paper.

**Subtask 3: analyze cancer-specific 5(h)mC biomarker using DNA from 90 prostate cancer tissues.**

To identify cancer-specific 5(h)mC biomarkers, we are in the process of sequencing more clinical samples in addition to Table 4. First, with the aid of the Northwestern Biorepository Network and Prostate SPORE, we have identified an additional sample set consisting of 20 prostate cancer tissues. This sample set contains 10 prostate cancer tissues who experienced subsequent progression of disease and 10 prostate cancer tissues, which did not progress, representing indolent disease. Using our highly optimized 5hmCSL-Seq assay, we will identify 5hmC alterations specifically associated with disease recurrence.

Second, our results indicate 5hmC alterations participate not only in the Benign to PCa transition, but also the transition from primary PCa to CRPC. Dr. Chuan He's group has recently demonstrated the 5hmCSL-Seq technique is accurate in the analysis of cell-free DNA (cfDNA) arising from cancer tissues. Therefore, we aim to use 5hmC as a diagnostic and prognostic tool for the monitoring of CRPC patients. In order to expedite this process we have develop a collaboration with Dr. Johann De Bono at The Institute of Cancer Research, London, who recently published research on cfDNA levels in CRPC patients (reference – PubMed ID: 28450425). They have performed comprehensive collection of CRPC patient cfDNA and have agreed to collaborate and share these samples. These work is still currently ongoing and is supported through institutional funds after this grant ended.

- **What opportunities for training and professional development has the project provided?**

***"Training" activities:***

Postdoctoral fellows Dr. Shangze Li, Bing Song, Nathan Damaschke, and graduate student Angela Yang, junior faculty Dr. Jonathan Zhao have received one-on-one training by discussion and meetings with the PIs and also by collaborating with the He's laboratory. They have gained extensive training in the study of DNA methylation and in various methylation assays. Dr. Damaschke, Yang, and Zhao have also obtained hands-on training the bioinformatics analysis of genomics data through the execution of the project.

At the University of Chicago, Mr. Xingyu Lu and Ms. Miao Yu (HHMI international pre-doctoral fellow) have been involved in the research. They have gained significant knowledge on prostate cancer research through interactions with the Yu laboratory. A new postdoc, Dr. Lulu Hu, has taken over the project and started to learn prostate cancer research.

***"Professional development" activities:***

The PI Dr. Yu has given lectures in various universities, including University of Michigan, Baylor College of Medicine, and Duke University on related topics.

Professor Chuan He has been given lectures on 5hmC in various scientific meetings and visits of other schools.

- **How were the results disseminated to communities of interest?**

*Nothing to Report*

- **What do you plan to do during the next reporting period to accomplish the goals?**

*Nothing to Report*

4. **IMPACT:**

- **What was the impact on the development of the principal discipline(s) of the project?**

*Nothing to Report*

- **What was the impact on other disciplines?**

*Nothing to Report*

- **What was the impact on technology transfer?**

*Nothing to Report*

- **What was the impact on society beyond science and technology?**

*Nothing to Report.*

5. **CHANGES/PROBLEMS:**

- **Changes in approach and reasons for change**

*Nothing to Report*

- **Actual or anticipated problems or delays and actions or plans to resolve them**

*Nothing to Report*

- **Changes that had a significant impact on expenditures**

*Nothing to Report*

- **Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents**

*Nothing to Report*

6. **PRODUCTS:**

- **Publications, conference papers, and presentations**

- **Journal publications.**

1) Yang YA, Zhao JC, Fong KW, Kim J, Li S, Song C, Song B, Zheng B, He C, Yu J. FOXA1 potentiates lineage-specific enhancer activation through modulating TET1 expression and function. **Nucleic Acids Res**. **2016** Jun 1. pii: gkw498. [Epub ahead of print]. Acknowledgement of federal support.
2) Yang YA, Kim J, Yu J. Influence of oncogenic transcription factors on chromatin conformation and implications in prostate cancer. **Appl Clin Genet**. **2014** May 12. PMID: 24876790. Acknowledgement of federal support.
3) Jin HJ, Zhao JC, Wu L, Kim J, Yu J. Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. **Nat Commun**. **2014** May 30. PMID: 24875621. Acknowledgement of federal support.
4) Xue XY, Yang YA, Zhang A, Fong KW, Kim J, Song B, Li S, Zhao JC, Yu J. LncRNA HOTAIR enhances ER signaling and confers tamoxifen resistance in breast cancer. **Oncogene**. **2015** Sep 14. PMID: 26364613 [Epub ahead of print] Acknowledgement of federal support.

- **Books or other non-periodical, one-time publications.**

*Nothing to Report*

- **Website(s) or other Internet site(s)**

*Nothing to Report*

- **Technologies or techniques**

*Nothing to report*

- **Inventions, patent applications, and/or licenses**

*Nothing to Report*

- **Other Products**

*Nothing to report*

## 7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

- **What individuals have worked on the project?**

| | |
|---|---|
| Name: | *Jindan Yu* |
| Project Role: | *PI* |
| Researcher Identifier (e.g. ORCID ID): | *1071255* |
| Nearest person month worked: | *1.2* |
| Contribution to Project: | *Dr. Yu supervised all the work related to this project, designed the experiments, facilitated collaboration between researchers, write reports and manuscripts.* |
| Funding Support: | |

-

| | |
|---|---|
| Name: | *Angela Yang* |
| Project Role: | *Graduate Student* |
| Researcher Identifier (e.g. ORCID ID): | *2654097* |
| Nearest person month worked: | *2.52* |
| Contribution to Project: | *Angela performed the work related to MeDIP and hMeDIP pull down assays, performed cell culture, DNA isolation, and methylation assays. She also performed analysis of the next-generation sequencing data by checking the results in genome-browser and examine how FOXA1 regulates TET1 and enhancer activation.* |
| Funding Support: | T32 Carcinogenesis training grant |

| | |
|---|---|
| Name: | *Shangze Li* |

| | |
|---|---|
| Project Role: | *Postdoctoral Fellow* |
| Researcher Identifier (e.g. ORCID ID): | *1074460* |
| Nearest person month worked: | *2.4* |
| Contribution to Project: | *Dr. Li has prepared all sequencing libraries for next-generation sequencing using the Illumina sample prep protocol and performed work in arranging and receiving the sequences.* |
| Funding Support: | |

| | |
|---|---|
| Name: | *Jonathan Zhao* |
| Project Role: | *Bioinformatician* |
| Researcher Identifier (e.g. ORCID ID): | *1082016* |
| Nearest person month worked: | *3.6* |
| Contribution to Project: | *Dr. Zhao is responsible for bioinformatics analysis of all next-generation sequencing data.* |
| Funding Support: | |

| | |
|---|---|
| Name: | *Nathan Damaschke* |
| Project Role: | *Postdoc* |
| Researcher Identifier (e.g. ORCID ID): | *1094643* |
| Nearest person month worked: | *3.00* |
| Contribution to Project: | *Dr. Damaschke is responsible for bioinformatics analysis of 5hmC-seq and MeDIP-seq data to identify cancer-specific methylation biomarkers.* |
| Funding Support: | |

- **Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?**

*Nothing to Report*

- **What other organizations were involved as partners?**

**Organization Name:** ***University of Chicago***
**Location of Organization:** *(if foreign location list country): Chicago, IL*
**Partner's contribution to the projec**t *(identify one or more): They performed the 5mCSL and 5hmCSL pull down involved in all experiments and arranged for part of the next-generation sequencing.*
**Collaboration** *(e.g., partner's staff work with project staff on the project)*; Mr. Xingyu Lu and Ms. Miao Yu at the University of Chicago participated in the 5mCSL and 5hmCSL pull down assays in prostate cancer DNAs that the Northwestern lab isolated and provided.
**Personnel exchanges** *(e.g., project staff and/or partner's staff use each other's facilities, work at each other's site)*; and

8. **SPECIAL REPORTING REQUIREMENTS**

**COLLABORATIVE AWARDS:** *NA.*
**QUAD CHARTS:** *NA.*

9. **APPENDICES:**

**Appendix 1:** Yang YA, Zhao JC, Fong KW, Kim J, Li S, Song C, Song B, Zheng B, He C, **Yu J**. *FOXA1 potentiates lineage-specific enhancer activation through modulating TET1 expression and function.* **Nucleic Acids Res.** 2016 Jun 1. pii: gkw498. [Epub ahead of print]. A*cknowledgement of federal support.*

**Appendix 2:** Yang YA, Kim J, **Yu J**. Influence of oncogenic transcription factors on chromatin conformation and implications in prostate cancer. ***Appl Clin Genet.*** 2014 May 12. PMID: 24876790. *acknowledgement of federal support.*

**Appendix 3:** Jin HJ, Zhao JC, Wu L, Kim J, **Yu J**. Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. ***Nat Commun.*** 2014 May 30. PMID: 24875621. *acknowledgement of federal support.*

**Appendix 4:** Xue XY, Yang YA, Zhang A, Fong KW, Kim J, Song B, Li S, Zhao JC, Yu J. LncRNA HOTAIR enhances ER signaling and confers tamoxifen resistance in breast cancer. Oncogene. 2015 Sep 14. PMID: 26364613 [Epub ahead of print] *acknowledgement of federal support.*

Nucleic Acids Research Advance Access published June 17, 2016

*Nucleic Acids Research, 2016* **1**
*doi: 10.1093/nar/gkw498*

# FOXA1 potentiates lineage-specific enhancer activation through modulating TET1 expression and function

**Yeqing A. Yang[1,†], Jonathan C. Zhao[1,†], Ka-wing Fong[1], Jung Kim[1], Shangze Li[1], Chunxiao Song[2], Bing Song[1], Bin Zheng[1], Chuan He[2,3] and Jindan Yu[1,4,*]**

[1]Division of Hematology/Oncology, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA, [2]Department of Chemistry, Department of Biochemistry and Molecular Biology, and Institute for Biophysical Dynamics, the University of Chicago, Chicago, IL 60637, USA, [3]Howard Hughes Medical Institute, the University of Chicago, Chicago, IL 60637, USA and [4]Robert H. Lurie Comprehensive Cancer Center, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA

Received September 29, 2015; Revised May 23, 2016; Accepted May 24, 2016



## ABSTRACT

**Forkhead box A1 (FOXA1) is an FKHD family protein that plays pioneering roles in lineage-specific enhancer activation and gene transcription. Through genome-wide location analyses, here we show that FOXA1 expression and occupancy are, in turn, required for the maintenance of these epigenetic signatures, namely DNA hypomethylation and histone 3 lysine 4 methylation. Mechanistically, this involves TET1, a 5-methylcytosine dioxygenase. We found that FOXA1 induces TET1 expression via direct binding to its cis-regulatory elements. Further, FOXA1 physically interacts with the TET1 protein through its CXXC domain. TET1 thus co-occupies FOXA1-dependent enhancers and mediates local DNA demethylation and concomitant histone 3 lysine 4 methylation, further potentiating FOXA1 recruitment. Consequently, FOXA1 binding events are markedly reduced following TET1 depletion. Together, our results suggest that FOXA1 is not only able to recognize but also remodel the epigenetic signatures at lineage-specific enhancers, which is mediated, at least in part, by a feed-forward regulatory loop between FOXA1 and TET1.**

## INTRODUCTION

Forkhead box A1 (FOXA1; also known as hepatocyte nuclear factor 3 α or HNF3A) belongs to the forkhead family of transcription factors and is known to play a pivotal role for the postnatal development of the mammary and prostate glands (1). FOXA1 is critical in directing hormone receptor-dependent transcriptional programs to regulate prostate- or breast-specific gene expression and cell differentiation (2,3). FOXA1 acts as a 'pioneer transcription factor' that can associate with compact chromatin to increase local chromatin accessibility and facilitate the recruitment of other transcription factors including nuclear receptors to these sites (4). Genome-wide location analyses have reported that FOXA1 preferentially recognizes and binds lineage-specific enhancers that are demarcated by active histone modifications including histone H3 lysine 4 mono- and di-methylation (H3K4me1, me2) (5), histone 27 acetylation (H3K27ac) (6), as well as local DNA hypomethylation (7). On the other hand, enforced expression of FOXA1 and its subsequent recruitment to enhancers lead to DNA demethylation and *de novo* gain of H3K4me1, suggesting that FOXA1 is able to remodel heterochromatic regions (7,8). However, the molecular mechanisms by which FOXA1 imposes this chromatin remodeling have not been characterized.

TET (ten-eleven translocation) proteins are a family of DNA hydroxylases that oxidize the methyl group at the C5 position of methylated cytosine, enzymatically converting 5-methylcytosine (5mC) into 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) in a sequential and iterative manner, ultimately leading to the removal of DNA methylation (9,10). Through catalyzing DNA demethylation, TET proteins play important roles in embryonic stem cell maintenance and in regulating appropriate lineage differentiation of these cells. These activities can be linked to the ability of DNA demethylation in modulating transcription factor occupancy and *vice versa* (11,12). During neural and adipocyte differentiation, dynamic hydroxmethylation has been associated with lineage-specific distal regulatory regions and represents an early

*To whom correspondence should be addressed. Tel: +1 312 503 1761; Fax: +1 312 503 0189; Email: jindan-yu@northwestern.edu
†These authors contributed equally to the paper as first authors.

event of enhancer activation (13). Concordantly, a separate study has demonstrated that deletion of Tet2 led to extensive loss of 5hmC and gain of DNA hypermethylation at enhancers and modulates enhancer activity of differentiation-related genes (14). However, the roles of TET proteins in FOXA1 recruitment and regulation of prostate lineage-specific enhancers are yet to be delineated.

Here, we show that TET1 is a direct target of FOXA1-mediated transcriptional activation. Further, TET1 physically interacts with the FOXA1 protein and modulates local DNA demethylation that in turn facilitates and stabilizes the recruitment of FOXA1. FOXA1 and TET1 thus form a feed-forward loop that activates lineage-specific enhancers. Not only does this mechanism provide a new perspective on the dynamic functional significance of the newly discovered TET1 DNA hydroxylase, but also offer insight into the molecular details underlying FOXA1's ability to fine-tune and modulate lineage-specific enhancer activation. As FOXA1 is a critical regulator and a top mutated gene in multiple cancers such as breast and prostate cancers (15), our study thus forms the framework for future understanding of the roles of TET1 in lineage-specific gene expression and cancer progression.

## MATERIALS AND METHODS

### Cell lines, plasmids and antibodies

Prostate cancer cell lines LNCaP, VCaP, 22Rv1, BPH1, RWPE-1, DU145 and human embryonic kidney cell line HEK293T cells were obtained from American Type Culture Collection and cultured in either RPMI1640 or Dulbecco's modified Eagle's medium with 10% fetal bovine serum (FBS). For FOXA1 and TET1 FL and domain constructs, human FOXA1 and TET1 cDNA were amplified by reverse transcription polymerase chain reaction (PCR) from LNCaP cells and pENTR223 TET1 (Harvard Plasmid), respectively, and cloned into the entry vector pCR8/GW/TOPO (Invitrogen). Adenoviral construct expressing FOXA1 was generated by recombining pCR8-FOXA1 with pAD/CMV/V5 using LR Clonase II (Invitrogen). Overexpression constructs for TET1 were generated by recombination of pCR8-TET1 with NTSFB destination vector or pLenti CMV/TO Puro DEST (Addgene plasmid 17 293). The pGIPZ lentiviral control and FOXA1 shRNAs were purchased from Open Biosystems. Sequences for scramble (5′-GCGCGCTTTGTAGGATTCG-3′) and TET1 (5′-GTGGAGAAGTGGACACAAA-3′) shRNA were kindly provided by Dr Debabrata Chakravarti (Northwestern University), and cloned into pLKO lentiviral vector.

The antibodies used in this study include anti-FOXA1 (ab23738) and anti-GAPDH (ab9385) from Abcam, anti-TET1 (GTX627420 and GTX124207) from GeneTex, anti-FLAG (F1804 and F7425) from Sigma, anti-c-Myc (sc-789x) from Santa Cruz, anti-HA (ab9110) from Abcam, anti-alpha Tubulin (sc-32293) from Santa Cruz, anti-5mC (BI-MECY-0100) from Eurogentec, anti-5hmC (39769) from Active Motif, anti-H3K4me2 (07-030) from Millipore, anti-H3K27ac (ab4279) from Abcam.

### Luciferase reporter assay

TET1 promoter and enhancer luciferase reporter assays were conducted according to the manual of Dual-Luciferase Reporter Assay System from Promega. Briefly, LNCaP cells were seeded in a 24-well plate and co-transfected with the Renilla expression plasmid pRL-TK and the reporter constructs for TET1 promoter and enhancer in pGL4 vector. Cells were infected with LacZ (control) or FOXA1 adenovirus for 48 h to assess the effect of FOXA1 overexpression on luciferase activity. Conversely, to look at FOXA1 depletion effect, lentiviral-transduced shCtrl and shFOXA1 LNCaP cells were used for co-transfection of reporter constructs. Luciferase activities were determined 48–72 h post-transfection and normalized against Renilla internal control values.

### Immunofluorescence staining

Cells were fixed with 4% formaldehyde for 15 min at RT and then permeabilized in 0.1% Triton X-100 for 15 min at RT. Cells were then washed by phosphate buffered saline (PBS) for three times, followed by incubation with 5% normal goat serum for 30 min at RT. Subsequently, cells were incubated with primary antibody, the anti-FLAG antibody (Sigma) and anti-TET1 (Genetex), for 2 h at RT. After washing three times with PBS, cells were incubated with secondary antibody, Alexa Fluor 488 and 594 goat anti-rabbit IgG (Invitrogen), for 1h at RT. Finally, cells were washed three times with PBS and mounted using Prolong Gold Antifade Reagent (Invitrogen).

### Co-immunoprecipitation

Nuclear proteins were extracted from 293T or LNCaP cells (details provided in Supplementary Methods). For S protein pulldown, nuclear extracts were incubated with 30 μl S-protein agarose beads (Millipore) for 3 h at 4°C. The beads/protein complex was then washed four times, and eluted with 30 μl 2× sodium dodecyl sulphate (SDS) sample buffer and subjected to western blot analysis. For LNCaP endogenous co-immunoprecipitation (co-IP), nuclear extracts were incubated with 2 μg antibodies, anti-FOXA1 (Abcam), anti-TET1 (Genetex) and anti-rabbit IgG (Santa Cruz) overnight at 4°C. Dynabeads Protein A (Life Technologies), 25 μl per immunoprecipitation (IP), were added the next day and incubated for 1 h at 4°C. Similarly, the beads/protein complex was washed four times, and eluted with 30 μl 2× SDS sample buffer and subjected to western blot analysis.

### Chromatin immunoprecipitation (ChIP)

Chromatin immunoprecipitation (ChIP) experiments were carried out as previously described (16). All primers (listed in Supplementary Data) were designed using Primer 3 (http://frodo.wi.mit.edu/primer3/), synthesized by Integrated DNA Technologies and used for SYBR Green based real-time PCR. ChIP-quantitative PCR enrichment of target loci was normalized to input DNA and reported as Enrichment over input ±SEM. ChIP DNA was prepared into

libraries according to standard protocols using Bioo Scientific's DNA Sample Kit (cat. no. 514101). Libraries were sequenced using Illumina Hi-Seq platforms. Sequence reads were aligned to the Human Reference Genome (assembly hg19) using Burrows–Wheeler alignment tool (bwa) version 0.6.1. New high-throughput data generated in this study have been deposited in GEO database under accession number GSE73363.

### Methylated DNA immunoprecipitation (MeDIP)

Methylated DNA immunoprecipitation (MeDIP) was performed as previously described (17). Total genomic DNA was extracted using QIAamp DNA Mini Kit (Qiagen) and sonicated to obtain fragments between 300 and 1000 bp. Dynabeads M-280 Sheep anti-Mouse IgG (Invitrogen) were incubated with an anti-5-methylcytidine antibody (BI-MECY_0100, Eurogentec, Fremont, CA, USA) overnight at 4°C. The following day, 4 μg of sheared DNA was denatured by boiling at 95°C for 10 min followed by rapid cooling on ice, and subsequently added to the beads/antibody complex. On day 3, the beads were washed three times with PBS + 0.05% Triton X-100 and eluted from beads by incubation at 65°C for 5 min in 150 μl elution buffer (TE + 1% SDS). Elution was repeated for a total of two times. Total eluates were treated with proteinase K and incubated at 50°C for 2 h. QIAquick PCR purification kit (Qiagen) was used to purify the eluted DNA, and lastly qPCR was used to determine the enrichment of target genomic regions using gene-specific primers (listed in Supplementary Data). Enrichment of target loci was normalized to input DNA and reported as Enrichment over input ±SEM.

### 5hmC chemical labeling (hMe-Seal)

5hmC labeling experiments were performed as previously described (18). Briefly, genomic DNA was fragmented to an average of 400 bp and was incubated with 50 mM HEPES buffer (pH 7.9), 25 mM MgCl2, 100 mM UDP-6-N3-Glc and 2 mM βGT for 1 h at 37°C. The labeled DNA was purified by the QIAquick Nucleotide Removal kit (Qiagen) and eluted in $H_2O$. The click chemistry was performed with the addition of 150 mM of disulfide-biotin, and the mixture was incubated for 2 h at 37°C. The labeled DNA fragments were then purified by the QIAquick Nucleotide Removal kit (Qiagen) and enriched by Dynabeads Streptavidin C1 (Invitrogen), and subsequently released by dithiothreitol (DTT) treatment. The enriched DNA fragments were first purified by Micro Bio-Spin 6 spin columns (Bio-Rad) followed by MinElute PCR Purification Kit (Qiagen).

## RESULTS

### FOXA1 expression contributes to lineage-specific enhancer activation

To determine the correlation between FOXA1 and active enhancer marks, we re-analyzed previously published FOXA1 (GSE37345), H3K4me2 and H3K27ac ChIP-seq data (GSE27823) (19,20) and confirmed that FOXA1 binding sites (FXBS) are indeed enriched for H3K4me2 and H3K27ac (Figure 1A: top). Further, we performed MeDIP for 5mC and chemical labeling of 5hmC followed by deep sequencing, namely MeDIP-seq and hMe-Seal-seq (18,21), respectively, to map their genomic landscapes in LNCaP cells which express FOXA1. Bioinformatic analysis revealed that FXBS are depleted of 5mC, but enriched for 5hmC, being consistent with previous reports (7). In addition, we found that this correlation was much weaker in two other prostate cell lines namely PrEC and PC-3M, wherein FOXA1 expression is low, suggesting that FOXA1 expression and occupancy might contribute to DNA demethylation at local chromatin (Supplementary Figure S1A–C). Since it has been previously suggested that transcription factor binding sites can demonstrate the low 5mC high 5hmC signature in embryonic stem cells (12), we looked at DNA methylation profiles in LNCaP cells for two other transcription factors CTCF and AR and observed similar patterns for 5mC and 5hmC (Supplementary Figure S2A–D). As a measure of negative control, genomic regions 20 kb downstream from the FOXA1 peaks, which will be referred to as non-peak sites throughout this paper, were examined for epigenetic signatures but did not exhibit any distinct pattern (Figure 1B: bottom).

To further elaborate on this, we depleted FOXA1 in LNCaP cells through lentiviral shRNA transduction (Supplementary Figure S1D) and performed pulldown and deep sequencing of 5mC and 5hmC. Interestingly, although the average intensity of 5mC around all FOXA1-occupied sites was not hugely affected upon FOXA1 depletion (Figure 1B), there was a significant decrease in 5hmC (Figure 1C), whereas no changes were seen in either 5mC or 5hmC for non-peak sites (Figure 1D and E). Concordantly, active enhancer marks H3K4me2 and H3K27ac were decreased around FXBS following FOXA1 knockdown, supporting reduced enhancer activities (Supplementary Figure S1E and F). To ensure the reliability of this genome-wide phenomenon, as well as to examine the changes with a more sensitive method, we performed MeDIP and hMe-Seal followed by qPCR for individual genes. Expectedly, 5hmC was greatly reduced across a number of FXBS (Figure 1F). On the other hand, despite the fact that 5mC showed no obvious change on a global scale, MeDIP-PCR revealed moderate increases in 5mC upon FOXA1 knockdown (Supplementary Figure S3). Taking into consideration that 5hmC abundance represents only ∼10% of 5mC in embryonic stem cells (9), it is reasonable to observe a more significant change in 5hmC rather than 5mC. It can be inferred from these results that FOXA1 may be functioning to alter DNA methylation specifically at regions where it occupies to achieve a demethylated state while accumulating 5hmC marks, thus potentiating enhancer activation.

### FOXA1 positively regulates TET1 gene expression.

As DNA demethylation has recently been shown catalyzed by the TET proteins, we next examined whether TET gene expression is associated with FOXA1. We first performed qRT-PCR analysis of FOXA1 and TET1 transcript across a panel of 12 prostate cell lines (Figure 2A and B). Interestingly, like FOXA1, TET1 is in general expressed at a much higher level in AR-positive prostate cancer cell lines such as C4-2B and VCaP cells than in AR-negative cells includ-

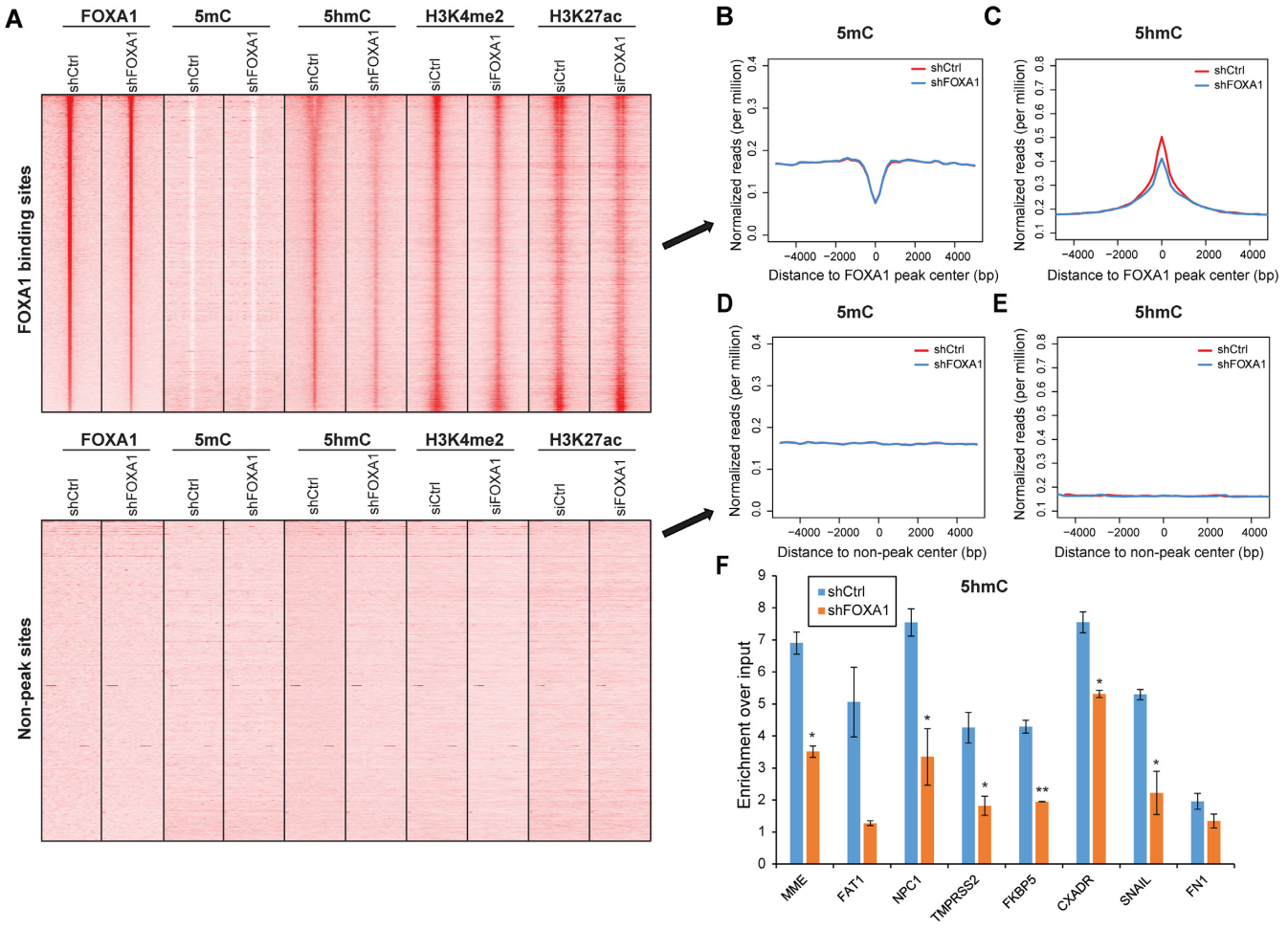



**Figure 1.** FOXA1 contributes to enhancer activation through epigenetic modifications. (**A**) Epigenetic signatures of FOXA1 binding sites (FXBS) (top panel) and non-peak sites taken 20 kb downstream (bottom panel) in control and shFOXA1 LNCaP cells. FOXA1 and H3K4me2/H3K27ac ChIP-seq data were obtained from publicly available datasets GSE37345 and GSE27823, respectively. Genomic landscapes of 5mC and 5hmC were determined by MeDIP and hMe-Seal, respectively, followed by deep sequencing. ChIP-seq read intensities of indicated epigenetic marks around (±5 kb) FXBS or non-peak regions in control (shCtrl) and FOXA1-knockdown (shFOXA1) cells were presented in heatmap format, ranked by read intensity of FOXA1 occupancy. (**B** and **C**) Average intensity plots of 5mC (B) and 5hmC (C) enrichment around all FXBS shown in A: top. (**D** and **E**) Average intensity plots of 5mC (D) and 5hmC (E) enrichment around all non-peak sites shown in A: bottom. (**F**) Locus-specific change in 5hmC by qPCR of hMe-Seal at representative FXBS for control and shFOXA1 LNCaP cells. Data shown is mean ± SEM of technical replicates from one representative experiment out of two. $*P < 0.05$ and $**P < 0.01$.

ing DU145 and RWPE. Further analysis showed that TET1 expression level is highly correlated ($r = 0.96$, $P < 0.001$) with that of FOXA1 (Supplementary Figure S4). This positively correlated expression between FOXA1 and TET1 was confirmed in three large prostate cancer patient datasets (Supplementary Figure S5A–C). As the correlation between FOXA1 and other TET proteins is relatively weaker, we decided to focus on TET1 in this study.

Since TET1 exhibited a similar expression pattern to FOXA1, we asked whether FOXA1 regulates TET1 gene expression. To test this, we first examined TET1 level in LNCaP cells with control or FOXA1 knockdown. Importantly, both TET1 transcript and protein levels were markedly decreased in LNCaP cells following FOXA1 knockdown (Figure 2C and D; Supplementary Figure S5D). Concordantly, depletion of FOXA1 in another independent prostate cancer cell lines C4-2B also resulted in a decrease in TET1 expression (Figure 2E). On the other hand, when FOXA1 was overexpressed in 22Rv1 cells through adenovirus infection, TET1 expression was augmented (Figure 2F), which was further validated in another prostate cancer cell line DU145 that contained low endogenous FOXA1 level (Figure 2G). To visualize the inductive effect of FOXA1 on TET1 at the cellular level, we performed immunofluorescence staining. TET1 was barely detectable in control DU145 cells infected with empty vector adenovirus (Figure 2H, top panel). However, upon infection with adenoviral FOXA1 (Flag-tagged, shown in red), TET1 staining (shown in green) was significantly enhanced (middle panel). Specifically, TET1 was stained positively in the majority of cells that had FOXA1 infection and overexpression, but not in the uninfected cells, as further illustrated in the zoomed-in microscopy images (Figure 2H, bottom panel). Taken together, our data support that FOXA1 positively regulates TET1 gene transcription.

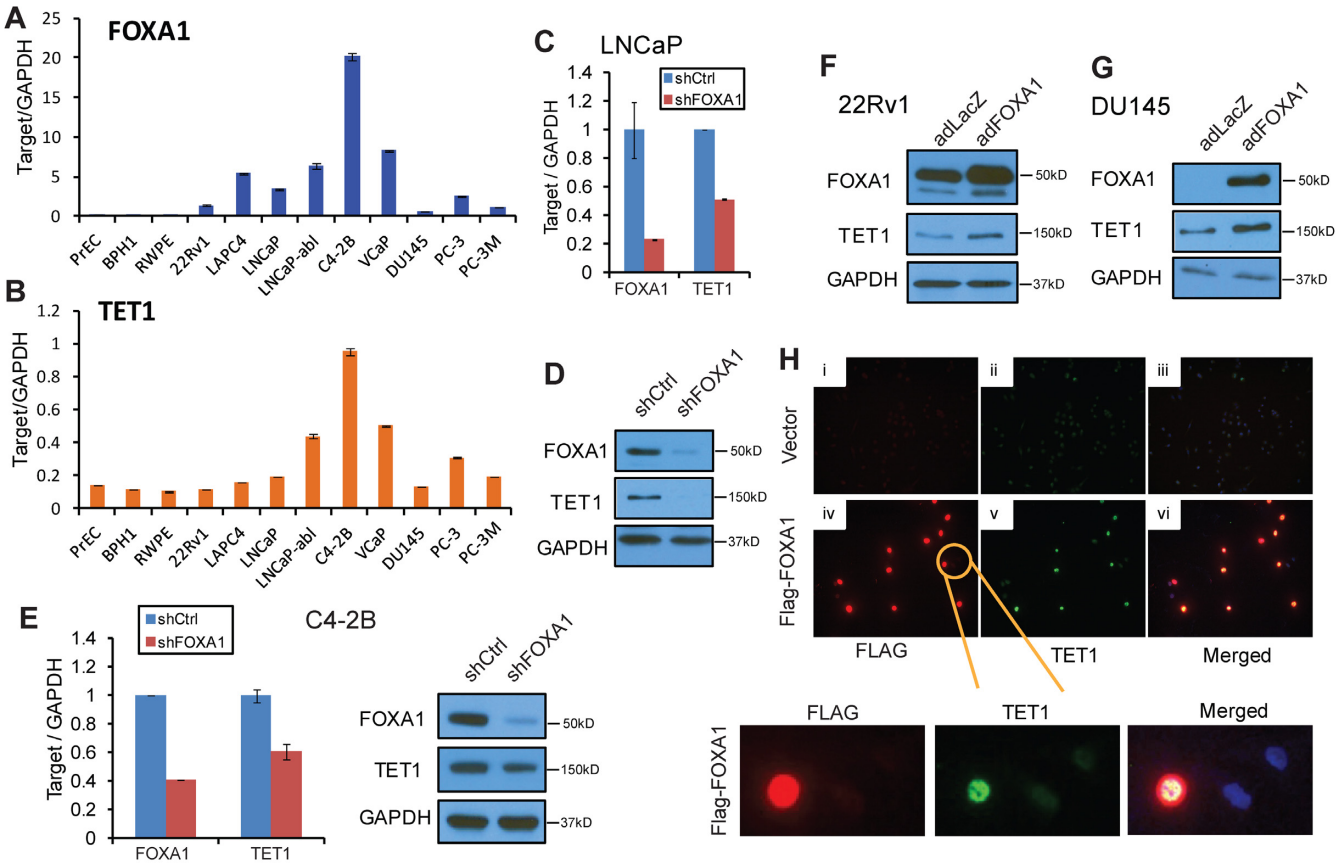



**Figure 2.** FOXA1 induces TET1 gene expression. (**A** and **B**) Correlated FOXA1 and TET1 gene expression in prostate cells. RNA was extracted from a panel of 12 prostate cell lines and analyzed by qRT-PCR for FOXA1 (A) and TET1 (B) gene expression. Data shown are mean ± SEM of technical replicates from one representative experiment out of three. (**C** and **D**) TET1 transcript (C) and protein (D) are downregulated following FOXA1 knockdown in LNCaP cells. LNCaP cells were infected with shCtrl or shFOXA1 lentivirus and subsequently subjected to qRT-PCR and western blot analysis. Data shown are one representative out of triplicate experiments. (**E**) TET1 is downregulated by FOXA1 knockdown in C4-2B cells. C4-2B cells were infected with shCtrl or shFOXA1 lentivirus for 8 h followed by puromycin selection for 4 days, and subsequently subjected to qRT-PCR and western blot analysis. Data shown are one representative out of triplicate experiments. (**F** and **G**) TET1 is upregulated following FOXA1 overexpression. The 22Rv1 (F) and DU145 (G) cells were infected with LacZ or FOXA1 adenovirus for 48 h and immunoblot was performed to assess FOXA1 and TET1 protein levels. (**H**) Positive TET1 staining in FOXA1-expressing cells. DU145 cells were infected with LacZ control (i–iii) or Flag-tagged FOXA1 (iv–vi) adenovirus for 48 h and then subjected to Immunofluorescence co-staining of FOXA1 and TET1. Bottom panel shows zoomed-in region containing both FOXA1-uninfected and -infected cells.

### TET1 is a direct transcriptional target of FOXA1

To determine how FOXA1 transcriptionally controls TET1 expression, we examined FOXA1 ChIP-seq data previously obtained from LNCaP cells (20), and observed a strong FOXA1 binding event within the intragenic region, between exons 3 and 4, of the TET1 gene (Figure 3A). Being consistent with FOXA1 as an enhancer regulator that modulates target genes through enhancer–promoter looping, we also found a weak FOXA1 binding event at the TET1 promoter. To validate the results of ChIP-seq, we performed ChIP-qPCR in LNCaP cells and found that FOXA1 is enriched at the TET1 enhancer for nearly 170-fold relative to IgG control, an enrichment level comparable to that at the Prostate-Specific Antigen (PSA, or KLK3) gene enhancer, and for about 10-fold at the TET 1 promoter (Figure 3B). A similarly strong enrichment of FOXA1 at the TET1 enhancer and promoter was also observed in an additional FOXA1-expressing cell line C4-2B (Supplementary Figure S6). Moreover, upon lentiviral knockdown, FOXA1 binding to its target site for the PSA gene was greatly diminished as expected, and similarly for TET1 enhancer and promoter, confirming that the ChIP enrichment signal was specific for FOXA1 (Figure 3C). Next, to examine whether FOXA1 occupancy at the TET1 enhancer and promoter leads to regulation of their transcriptional activities, we cloned these regions into reporter constructs. Luciferase assays showed that FOXA1 overexpression indeed significantly increased, whereas FOXA1 knockdown decreased, TET1 enhancer and promoter activities (Figure 3D and E). To further demonstrate that this regulation is due to FOXA1 occupancy at the TET1 enhancer and promoter, we analyzed the DNA sequences around the FOXA1 binding peaks for FKHD motifs within the TET1 enhancer as well as promoter. Through mutagenesis assays, we generated TET1 enhancer and promoter constructs with mutations to highly conserved FKHD motifs (Figure 3A, bottom panels). Importantly, luciferase assays revealed that mutations to the FKHD motifs abolished FOXA1 regulation of TET1

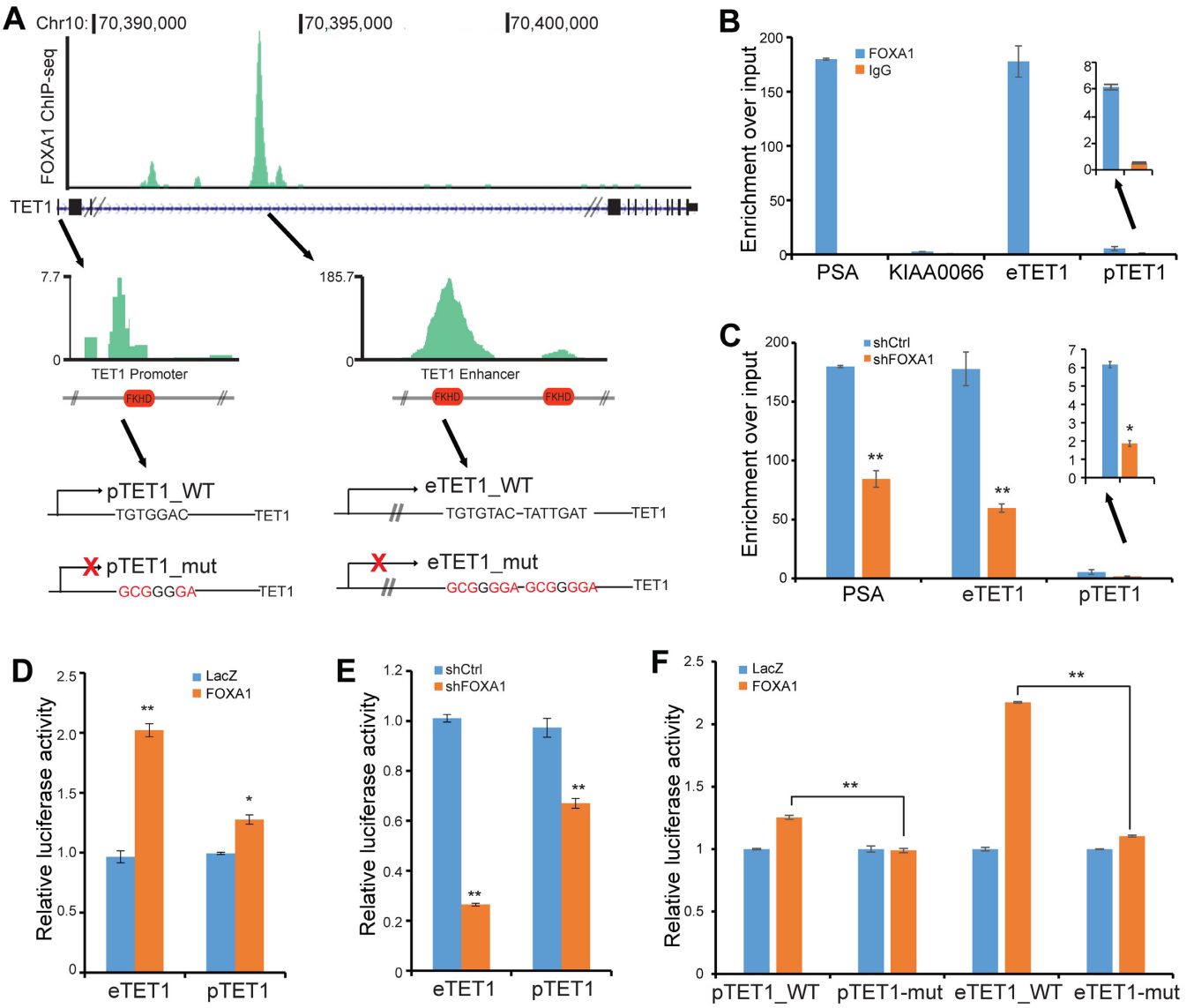



**Figure 3.** TET1 is a direct transcriptional target of FOXA1. (**A**) ChIP-seq showing FOXA1 binding events at TET1 promoter and enhancer. FOXA1 ChIP-seq was conducted in LNCaP cells and FOXA1 binding events were identified by HOMER and visualized in UCSC Genome Browser. FKHD motifs (indicated by red box) near FXBS were determined by JASPAR. DNA fragments containing FXBS at the TET1 promoter (pTET1) and enhancer (eTET1) were each cloned into pGL4 luciferase reporter construct with wild-type (WT) or mutated (mut) FKHD motif (mutated nt shown in red at the bottom panel). (**B**) ChIP-PCR validation of FOXA1 binding to TET1 enhancer and promoter in LNCaP cells. ChIP was performed using anti-FOXA1 and anti-IgG antibodies in LNCaP cells. ChIP-qPCR was performed using primers flanking the FOXA1 binding peaks at the TET1 enhancer (eTET1) and promoter (pTET). PSA is used as a positive control while KIAA0066 a negative control. Data shown are mean ± SEM of technical replicates from one representative experiment out of three. (**C**) FOXA1 occupancy at TET1 promoter and enhancer was decreased by FOXA1 knockdown. ChIP-qPCR using anti-FOXA1 antibody was carried out in control and FOXA1-depleted LNCaP cells. Data shown are mean ± SEM of technical replicates from one representative experiment out of three. $*P < 0.05$ and $**P < 0.01$. (**D** and **E**) FOXA1 positively regulates TET1 enhancer and promoter activities. TET1 enhancer and promoter reporter constructs were transfected into LNCaP cells with control or FOXA1 overexpression (D) and LNCaP cells with control or FOXA1 knockdown (E) for 48 h. Luciferase activities were determined and normalized to internal control Renilla reporter. Data shown are mean ± SEM of two independent experiments. $*P < 0.05$ and $**P < 0.01$. (**F**) FKHD motif is required for FOXA1-induced TET1 promoter and enhancer luciferase activities. Control and FOXA1-overexpressing LNCaP cells were transfected with either WT or mutated (depicted in A) TET1 promoter and enhancer reporter constructs. Luciferase activities were determined and normalized to internal control Renilla reporter. Data shown are mean ± SEM of two independent experiments. $*P < 0.05$ and $**P < 0.01$.

enhancer as well as promoter activities (Figure 3F). Taken together, our data support that FOXA1 directly binds to the regulatory elements of TET1 gene to induce its transcription. As FOXA1 contributes to local DNA demethylation (Figure 1) and TET1 is a known DNA demethylase, we hypothesized that TET1 may be attributable for DNA demethylation around the FXBS. To test this hypothesis, we started out by examining potential interactions between the FOXA1 and TET1 proteins.

### FOXA1 and TET1 proteins physically interact

By use of overexpression systems in 293T cells, we conducted co-IP experiments to assess whether physical interaction is present between ectopic FOXA1 and TET1 proteins. The 293T cells were co-transfected with Flag-tagged TET1 along with FOXA1 or empty vector. Successful expression of the ectopic proteins was confirmed by western blot analysis of the input lysate. IP using an anti-FOXA1 antibody followed by immunoblotting confirmed successful pulldown of FOXA1 itself as well as the TET1 protein, the latter only in the cells expressing both TET1 and FOXA1 (Figure 4A). To demonstrate the interaction through reversal co-IP, we cloned TET1 into the SFB-tagged expression vector, which enabled pulldown of the TET1 protein using S-protein agarose beads and detection by anti-Flag antibodies (22). Either SFB-vector control or SFB-TET1 was co-transfected with FOXA1 into 293T cells and their expression was confirmed by western blot analysis of the input lysate. S-protein pulldown followed by western blotting using anti-Flag validated successful enrichment of SFB-tag only or SFB-TET1 (of different sizes) in the corresponding lysates, while immunoblotting using anti-FOXA1 revealed FOXA1 pulldown only in the SFB-TET1-expressing cells (Figure 4B), supporting physical interaction between ectopic FOXA1 and TET1 proteins.

Next, we attempted to confirm this interaction between endogenous FOXA1 and TET1 proteins. LNCaP cell nuclear lysate was subjected to IP using rabbit anti-TET1, anti-FOXA1 and IgG control followed by western blotting with mouse anti-TET1 or anti-FOXA1 antibodies. Our results demonstrated that TET1 and FOXA1 antibodies are able to pull down each other, supporting strong protein interactions (Figure 4C). To address the potential involvement of DNA in mediating this interaction, we performed co-IP in the presence or absence of ethidium bromide. Notably, our results demonstrated persistent interaction between FOXA1 and TET1 proteins in the presence of ethidium bromide, thus indicating that DNA was not required for their association (Supplementary Figure S7A). Moreover, this interaction between endogenous FOXA1 and TET1 proteins was also confirmed in C4-2B cells (Supplementary Figure S7B).

To further determine which domains of the TET1 protein are important for its interaction with FOXA1, we generated four Myc-tagged TET1 domain constructs, namely the N-terminal, CXXC, middle and CD domains, which were co-transfected with SFB-tagged FOXA1 into 293T cells. S-protein pulldown followed by western blot analysis showed that only the TET1 fragment containing the CXXC module was able to bind FOXA1 (Figure 4D). On the other hand, we attempted to map out the FOXA1 domain that is responsible for its interaction with the TET1 protein. Similarly, we created three Flag-tagged FOXA1 domain constructs, namely N-terminal, Forkhead (FH) and C-terminal domains, which were co-transfected with SFB-tagged TET1-CXXC domain into 293T cells. Western blot analysis confirmed the expression of various FOXA1 domains of different sizes as expected (Figure 4E). S-protein pulldown of TET1 followed by western blotting revealed that only the FH-containing domain of FOXA1 protein is able to interact with the TET1-CXXC domain. Moreover, we also performed *in vitro* pulldown assay utilizing purified TET1-CXXC and FOXA1-FH domain proteins, which confirmed that the two proteins directly interact (Figure 4F). As the CXXC zinc finger module in Tet3 protein has been shown critical for specific chromatin targeting, while its enzymatic domain modulates its biological function (23), we hypothesized that TET1 interaction with FOXA1 through its CXXC domain may be important for its recruitment to FXBS where it carries out hydroxylation on methylated CpG's closeby through its CD domain. Therefore, we next asked whether TET1 regulates DNA demethylation and alters epigenetic modifications around FXBS.

### TET1 mediates active epigenetic modification at FOXA1-dependent enhancers

To determine whether TET1 affects the epigenetic environment at FOXA1-occupied enhancers, we first tested whether TET1 is able to co-occupy FOXA1-bound genomic regions. As human anti-TET1 antibody has not been well-established for ChIP, we transfected HA-tagged TET1 into LNCaP prostate cancer cells, validated by western blot in Supplementary Figure S8 and performed ChIP using ChIP-grade anti-HA antibody. ChIP-qPCR confirmed much stronger HA (TET1) enrichment at FXBS in cells expressing HA-TET1 than cells transfected with HA-control vector (Figure 5A). Next, to examine how TET1 alters DNA methylation around these FOXA1-bound regions, we performed TET1 knockdown using shRNA (Figure 5B). Western blots of different exposure times were included to show that TET1 was detected much more strongly at 150 kD, while also giving a very weak band above 250 kD, both of which were depleted upon shRNA knockdown (Supplementary Figure S5D). As TET1 is a DNA demethylase that catalyzes 5mC–5hmC, we next sought to determine the level of 5hmC and 5mC in TET1-knockdown cells. Dot blot experiment confirmed significant reduction of total 5hmC abundance in shTET1 cells (Supplementary Figure S9A). Further, hMe-Seal-seq revealed a remarkable decrease of total 5hmC-enriched regions following TET1 knockdown (Figure 5C). By contrast, 5mC as measured by MeDIP-seq was increased nearly 33% (Figure 5D). Average intensity view of all peaks showed that hMe-Seal signals were significantly decreased, while MeDIP signals increased upon TET1-knockdown (Figure 5E and Supplementary Figure S9B). Focused analysis of these epigenetic modifications around FXBS confirmed an overall decrease of 5hmC and increase of 5mC following TET1 depletion, suggesting that TET1 is critical for the maintenance of the demethylated state of these enhancers (Figure 5F and G). As DNA methy-

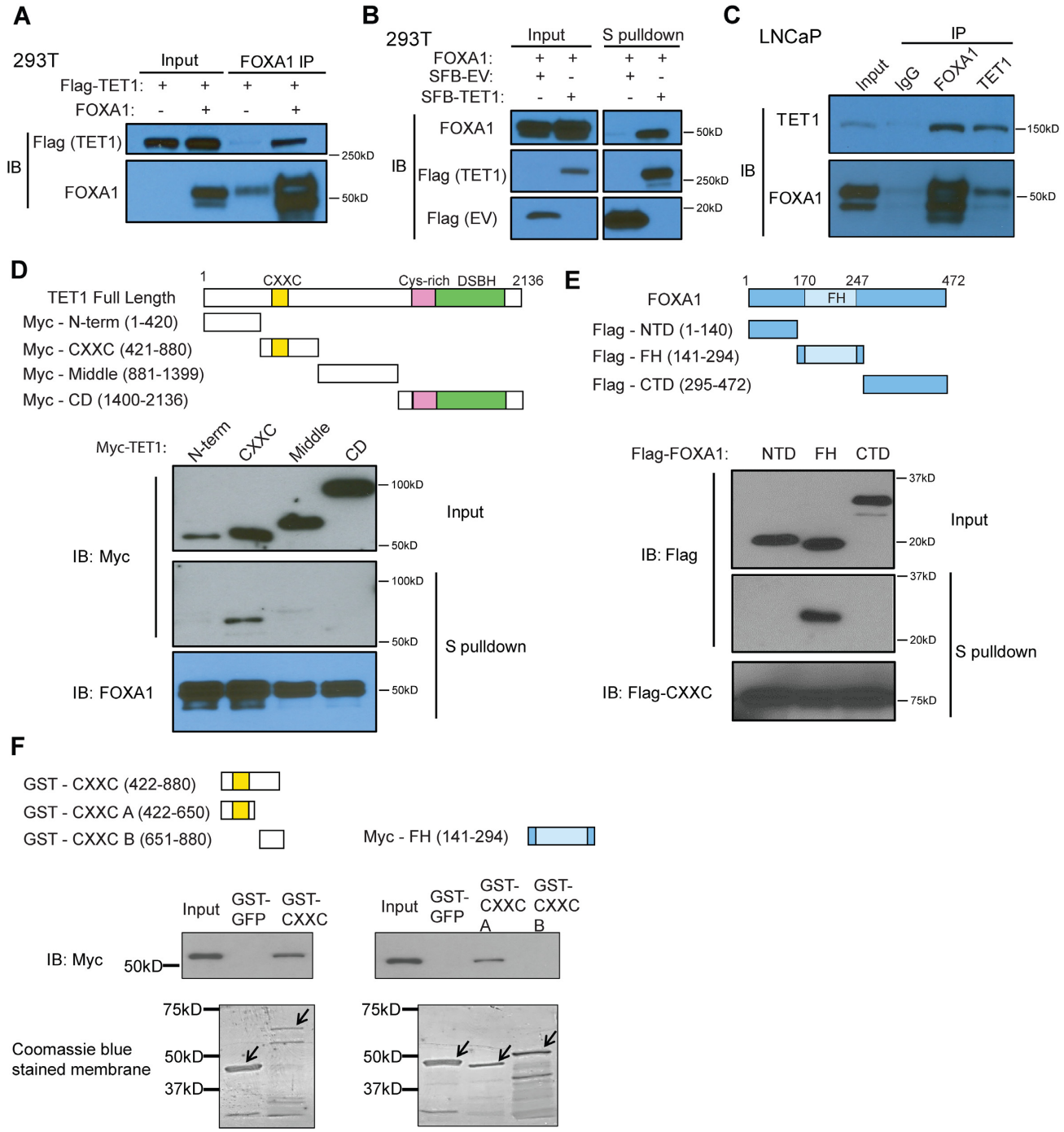



**Figure 4.** FOXA1 and TET1 proteins physically interact. (**A**) Immunoprecipitation of ectopic FOXA1 pulled down TET1 protein. The 293T cells were transfected with Flag-TET1, either alone or together with FOXA1, for 48 h and then subjected to immunoprecipitation using an FOXA1 antibody. Whole cell (Input) and IP-enriched lysates were then analyzed by western blotting using anti-Flag (TET1) and anti-FOXA1 antibodies. (**B**) Ectopic TET1 immunoprecipitation pulled down FOXA1 protein. The 293T cells were co-transfected with FOXA1 and SFB-tagged empty vector (EV) or TET1 for 48 h before immunoprecipitation using S beads, which will pull down SFB-EV or SFB-TET1. The input and IP-enriched cell lysates were then subjected to western blotting using anti-FOXA1 and anti-Flag (for SFB-EV or SFB-TET1) antibodies. (**C**) Endogenous FOXA1 and TET1 proteins interact in LNCaP cells. LNCaP cells were subjected to immunoprecipitation using anti-FOXA1, anti-TET1 and IgG control, followed by western blotting of FOXA1 and TET1 proteins. (**D**) TET1 CXXC domain interacts with the FOXA1 protein. 293T cells were co-transfected with SFB-FOXA1 along with various Myc-tagged TET1 domain constructs. The expression of TET1 domains in whole cell lysate (input) was confirmed by western blotting using anti-Myc. Cell lysates were then subjected to S pull down (of FOXA1) and subsequently western blot analysis using anti-FOXA1 and anti-Myc antibodies. (**E**) FOXA1 FH (Forkhead-containing) domain interacts with TET1 CXXC domain. 293T cells were co-transfected with SFB-CXXC along with various Flag-tagged FOXA1 domain constructs and subjected to S pull down (of TET1-CXXC) followed by western blotting using an anti-Flag antibody. (**F**) *Invitro* interaction assay was conducted using purified proteins of TET1 CXXC domain and FOXA1 Forkhead domain. CXXC domain was tagged with GST and further subdivided into fragments A and B (the 'C-X-X-C' motif was located in residues 590–609 in fragment A), and FH domain was tagged with Myc. Arrows point to expression of proteins according to their expected size.

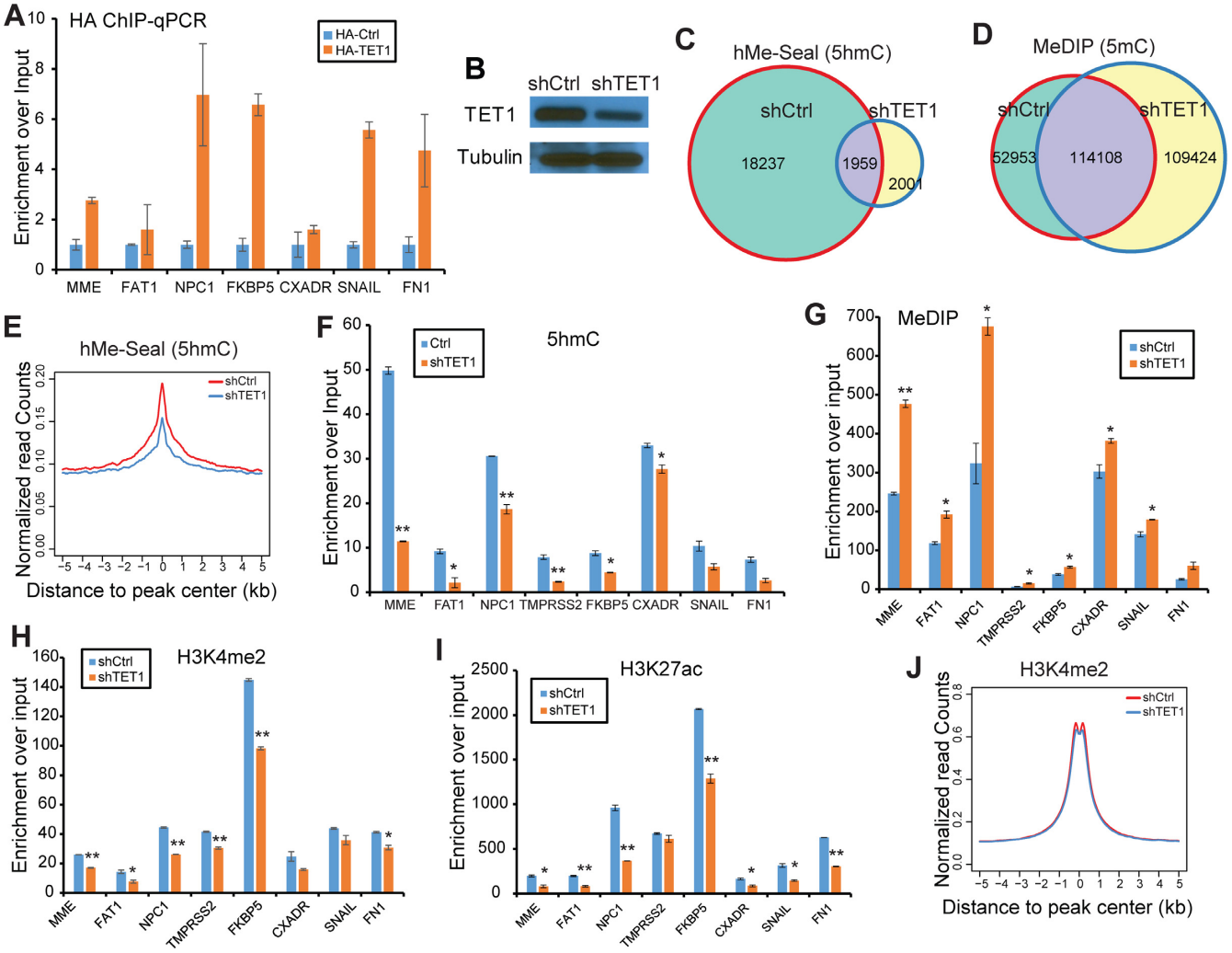



**Figure 5.** TET1 mediates active epigenetic modification at FOXA1-bound enhancers. (**A**) TET1 co-occupies FXBS. LNCaP cells were transfected with HA-tagged empty vector or TET1 constructs and were subsequently used for ChIP with anti-HA antibody. HA ChIP-qPCR was performed using primers flanking a number of FXBS. Data shown are mean ± SEM of technical replicates from one representative experiment out of three. (**B**) Western blots confirming TET1 knockdown. LNCaP cells were infected with either scramble or shTET1 lentivirus followed by puromycin selection for 4 days before western blot analysis. Tubulin is used as a loading control. (**C** and **D**) Venn Diagrams showing alterations in global genomic regions enriched for 5hmC (C) and 5mC (D) following TET1 knockdown. LNCaP cells with control or shTET1 were subjected to hMe-Seal-seq and MeDIP-seq for genome-wide location analysis of 5hmC and 5mC, respectively, which were subsequently compared between control and TET1-depleted cells. (**E**) Average intensity plot of normalized hMe-Seal-seq reads around (±5 kb) FXBS. (**F**–**I**) TET1 knockdown led to altered epigenetic signatures at FXBS. LNCaP cells with control or TET1 knockdown were subjected to hMe-Seal (F) and MeDIP (G) and ChIP using anti-H4K4me2 (H) and anti-H3K27ac (I) antibodies, followed by qPCR analysis with site-specific primers. Data shown are mean ± SEM of technical replicates from one representative experiment out of two. $*P < 0.05$ and $**P < 0.01$. (**J**) Average intensity plots of normalized H3K4me2-seq reads around (±5 kb) FXBS.

lation has been shown to inhibit enhancer activation (7), we next asked whether TET1 knockdown prohibits enhancer activation at FXBS. ChIP-qPCR showed that indeed H3K4me2 and H3K27ac were both significantly reduced following TET1 depletion (Figure 5H and I). ChIP-seq further confirmed a global decrease of H3K4me2 level in TET1-knockdown (Figure 5J). Taken together, our data support that TET1 expression contributes to the activation of FOXA1-target enhancers through mediating active DNA demethylation.

## TET1 expression is required for FOXA1 recruitment to target enhancers

Since it has been reported that DNA methylation and removal of H3K4me2 could impair FOXA1 binding (5,7), the changes in DNA methylation and histone modification events observed following TET1 depletion were suggestive of disrupted FOXA1 recruitment to these regions. To test this, we performed FOXA1 ChIP-seq in control and TET1-knockdown LNCaP cells to determine whether TET1 depletion is able to regulate FOXA1 chromatin targeting. A global assessment of the total binding events before and after TET1 knockdown demonstrated that a significant proportion of FOXA1 binding events were lost upon

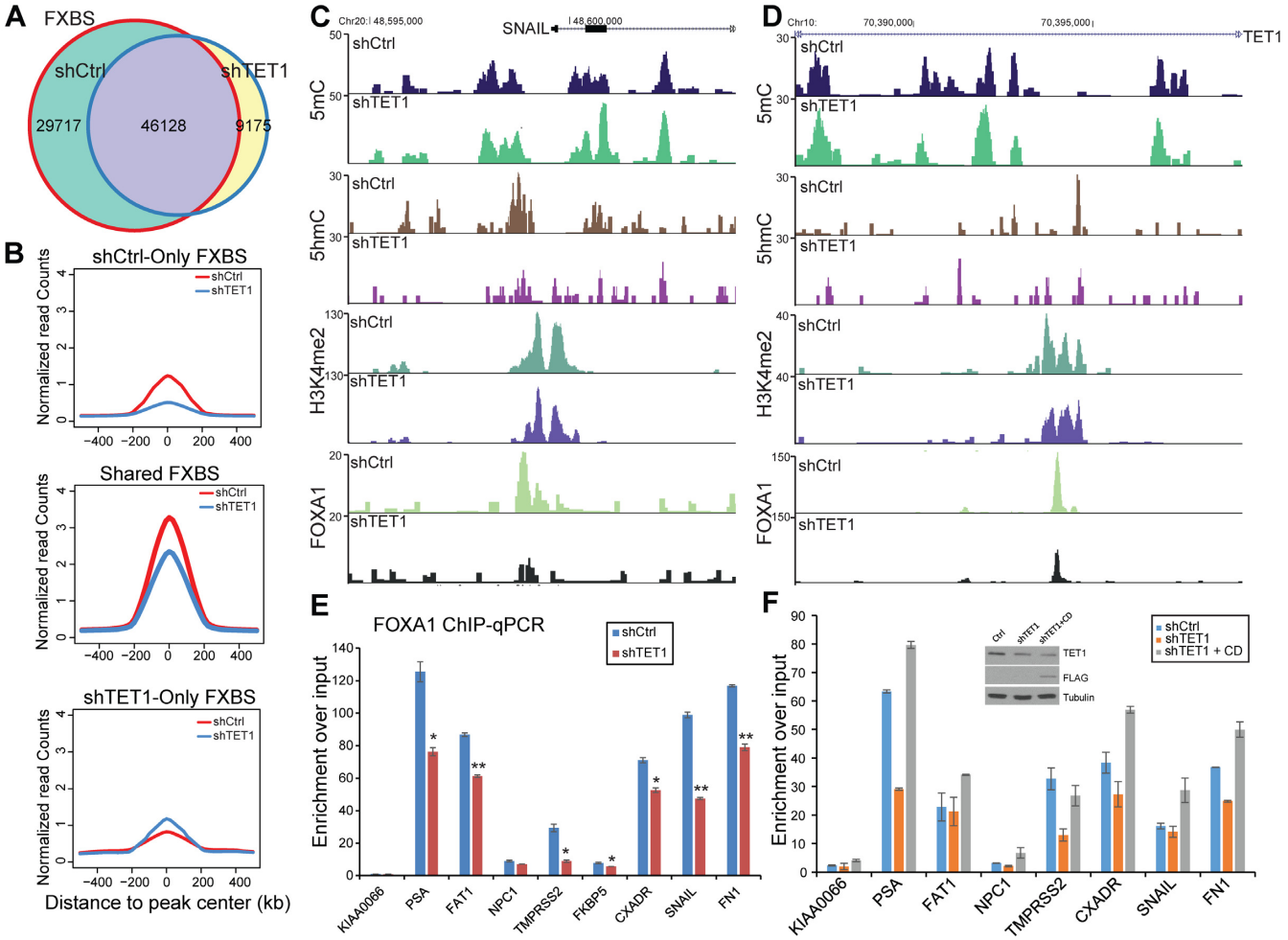



**Figure 6.** TET1 is required for FOXA1 recruitment to lineage-specific enhancers. (**A**) Venn diagram showing overlap of FXBS in control and FOXA1-knockdown LNCaP cells. (**B**) Average FOXA1 ChIP-seq read intensity around (±500 bp) shCtrl-only, shared and shTET1-only FXBS identified from overlap Venn diagram in (A). (**C** and **D**) Genome browser views of epigenetic modifications at the regulatory regions of FOXA1-target genes SNAIL (C) and TET1 itself (D). MeDIP-seq (5mC), hMe-Seal-seq (5hmC), H3K4me2 and FOXA1 ChIP-seq were performed in control and TET1-knockdown LNCaP cells. For each mark, the shCtrl and shTET1 tracks are shown on the same scale (Y-axis) for visual comparison of enrichment. (**E**) TET1 depletion attenuates FOXA1 recruitment to target enhancers. ChIP-qPCR was performed in control and shTET1 LNCaP cells using anti-FOXA1 antibody. Data shown are mean ± SEM of technical replicates from one representative out of triplicate experiments. $*P < 0.05$ and $**P < 0.01$. (**F**) Impaired FOXA1 recruitment in TET1-depleted cells is restored by TET1 CD overexpression. LNCaP cells were subjected to control or TET1 knockdown with or without concomitant TET1 CD overexpression. TET1 knockdown and CD domain (Flag-CD) overexpression were confirmed by western blot analysis (inset). Cells were subsequently used for ChIP with an anti-FOXA1 antibody followed by qPCR analysis. Data shown are mean ± SEM of technical replicates from one representative of duplicate experiments.

TET1 depletion (Figure 6A). The total number of FXBS was decreased from 76 000 to 55 000. In addition, the average intensity of FOXA1 binding events appeared to be much weaker even for the sites that were not fully abolished (i.e. shared sites) following TET1 knockdown (Figure 6B). Genome browser view of several FOXA1-dependent enhancers further illustrated significant loss of FOXA1 occupancy in TET1-depleted cells (Figure 6C and D; Supplementary Figure S10A and B). Meanwhile, DNA methylation at these enhancers was increased as indicated by enhanced 5mC but reduced 5hmC signals, while active enhancer mark H3K4me2 was decreased, being concordant with the genome-wide switch to repressive chromatin state. Moreover, ChIP-qPCR confirmed that TET1 knockdown significantly decreased FOXA1 occupancy at multiple target enhancers (Figure 6E).

As TET1 interacts with the FOXA1 protein through its CXXC domain but is known to carry out enzymatic activities through its CD domain, we next attempted to understand mechanistically whether CD-mediated DNA demethylation is sufficient to facilitate FOXA1 recruitment to target enhancers. A recent study has reported an interesting and important observation that the CD domain of TET proteins induces massive global DNA demethylation, whereas the function of full-length TET1 is much restricted to unmethylated CpG islands (24). We thus predict that CD domain may be able to restore FOXA1 recruitment in TET1-knockdown cells. To test this, we overexpressed the CD domain in LNCaP cells with TET1 knockdown. ChIP-

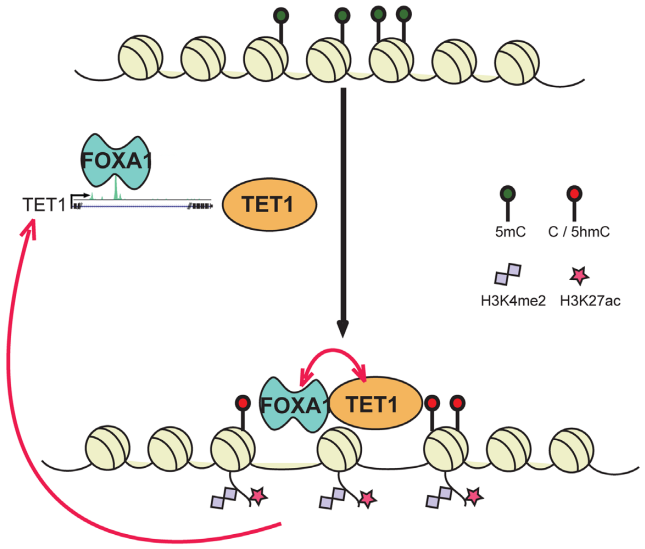



**Figure 7.** Schematic model depicting feed-forward regulation between FOXA1 and TET1 in lineage-specific enhancer activation. FOXA1 protein occupies at an intragenic enhancer of the TET1 gene to induce TET1 expression. Through direct interaction with FOXA1 protein, TET1 modulates DNA demethylation and subsequently H3K4 methylation and H3K27 acetylation at FOXA1-target enhancers, which in turn facilitates FOXA1 recruitment. Thus, FOXA1 and TET1 form a positive feedback loop in lineage-specific enhancer activation. FOXA1 is not only capable of recognizing but also modifying epigenetic signatures at lineage-specific enhancers.

qPCR confirmed that FOXA1 binding at target enhancers was decreased by TET1 knockdown, which, importantly, can be fully rescued by concomitant CD domain overexpression (Figure 6F). Taken together, our data support that TET1 facilitates FOXA1 recruitment to target enhancers through active demethylation.

## DISCUSSION

FOXA1 is a critical regulator of hormone-mediated gene expression in prostate and breast cancers. Much efforts have been devoted to understand the molecular basis for FOXA1's activity as a pioneer factor and studies in the past two decades have helped to paint a clearer picture of how FOXA1 activity is dependent on a number of epigenetic signatures that exhibit lineage specificity (5). Although FOXA1 has been shown to impose some effects on the epigenetic signatures around target enhancers (5), the molecular mechanisms by which FOXA1 remodels heterochromatin remain largely unknown. In the present study, we show that FOXA1 is able to directly regulate the transcription of *TET1* gene. Further, FOXA1 physically interacts with the TET1 protein, leading to DNA demethylation and H3K4me2/H3K27ac modifications at FOXA1-target enhancers. These changes in the epigenetic environment on the other hand enhance FOXA1 recruitment. Therefore, our data support a model wherein FOXA1 is not only able to recognize and bind enhancer regions, but contributes to *de novo* gain of H3K4 methylation and enhancer activation. The latter is mediated by, at least in part, a feed-forward loop between FOXA1 and TET1 where FOXA1 induces TET1 expression and binding at lineage-specific enhancers, which in turn facilitates and stabilizes FOXA1 recruitment through catalyzing DNA demethylation (Figure 7). Accompanying changes in DNA methylation are also reductions in H3K4me2 and H3K27ac upon FOXA1 depletion. Whether these are secondary to DNA demethylation or FOXA1/TET1 may regulate histone methyltransferases such as MLL are areas for future investigation.

TET1 has been implicated in the regulation of enhancer activation and lineage differentiation through DNA demethylation (13,14), the underlying mechanism of which, however, remains elusive. In this report, using prostate cancer cells as a model system, we demonstrated that TET1 contributes to FOXA1 recruitment to prostate-specific enhancers by modulating local epigenetic switch. In future studies, it will be interesting to investigate and compare how TET1 regulates epigenetic marks and FOXA1 recruitment in breast cancer, since FOXA1 has been shown to bind distinct, lineage-specific enhancers in prostate and breast cells (5). In addition, this study will pave the way to further investigation of how TET1, through modulation of FOXA1-dependent enhancer activation, regulates hormone-dependent gene expression and prostate and breast cancer progression.

The CXXC domain of TET proteins has been shown critical for specific chromatin targeting, while the CD domain modulates its enzymatic activity (23). Further, a recent study has reported that the full-length TET1 protein preferably binds to unmethylated CpG islands through its CXXC domain (24). Being consistent with these reports, we found that FOXA1 interacts with TET1 protein through its CXXC domain. Such interaction may be critical for targeting TET1 to prostate-specific enhancers denoted by FOXA1, which may be interesting lines for further investigation utilizing various TET1 deletion constructs and ChIP-seq experiments. Moreover, TET1 might similarly interact with other lineage-defining transcription factors and get recruited to distinct, lineage-specific enhancers in different cell types. By contrast, overexpression of the TET1 CD domain alone has been shown to induce massive global DNA demethylation (24). Indeed, in our study we found overexpression of CD domain is able to rescue the effects of TET1 knockdown on FOXA1 recruitment to target enhancers. Therefore, through interaction with other transcription factors, TET1 achieves its specificity to bind selected enhancers, where it carries out its role in the maintenance of hypomethylated landscape and regulation of lineage differentiation.

In conclusion, FOXA1 is a multipotent pioneer transcription factor, which is impressively capable of chromatin remodeling through not only histone displacement but also DNA demethylation by employing the DNA hydroxylase TET1. Collectively, through regulation of TET1 expression and function, FOXA1 is able to control the epigenetic signatures present at its cis-regulatory elements through a feed-forward loop, ultimately giving rise to chromatin relaxation and enhancer activation.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We would like to thank Dr Debabrata Chakravarti for providing the shTET1 construct.
*Author contributions*: Y.A.Y, K.F., J.K., S.L., C.S., B.S. and B.Z. performed the experiments. J.C.Z performed all bioinformatics analysis. J.Y. conceived and supervised the project. J.Y. and Y.A.Y. designed the experiments. C.H. supervised the MeDIP-seq and hMe-Seal-seq experiments. J.Y., Y.A.Y. and J.C.Z generated the figures and wrote the manuscript. All authors discussed the results, commented on the manuscript and declare no conflicts of interest.

## FUNDING

U.S. Department of Defense [W81XWH-13-1-0319 to J.Y., in part]; American Cancer Society Research Scholar Award [RSG-12-085-01 to J.Y.]; National Institutes of Health [R01HG006827 to C.H.]; Howard Hughes Medical Institute (to C.H.); NIH/NCI training grant [T32 CA09560 to Y.A.Y., in part]; NIH Training Program in Oncogenesis and Developmental Biology [T32 CA080621 to J.K., in part]. Funding for open access charge: National Institutes of Health [R01-CA172384 to J.Y.].
*Conflict of interest statement.* None declared.

## REFERENCES

1. Bernardo,G.M. and Keri,R.A. (2012) FOXA1: a transcription factor with parallel functions in development and cancer. *Biosci. Rep.*, **32**, 113–130.
2. Gao,N., Ishii,K., Mirosevich,J., Kuwajima,S., Oppenheimer,S.R., Roberts,R.L., Jiang,M., Yu,X., Shappell,S.B., Caprioli,R.M. *et al.* (2005) Forkhead box A1 regulates prostate ductal morphogenesis and promotes epithelial cell maturation. *Development*, **132**, 3431–3443.
3. Bernardo,G.M., Lozada,K.L., Miedler,J.D., Harburg,G., Hewitt,S.C., Mosley,J.D., Godwin,A.K., Korach,K.S., Visvader,J.E., Kaestner,K.H. *et al.* (2010) FOXA1 is an essential determinant of ERalpha expression and mammary ductal morphogenesis. *Development*, **137**, 2045–2054.
4. Jozwik,K.M. and Carroll,J.S. (2012) Pioneer factors in hormone-dependent cancers. *Nat. Rev. Cancer*, **12**, 381–385.
5. Lupien,M., Eeckhoute,J., Meyer,C.A., Wang,Q., Zhang,Y., Li,W., Carroll,J.S., Liu,X.S. and Brown,M. (2008) FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell*, **132**, 958–970.
6. Wang,J., Zhuang,J., Iyer,S., Lin,X., Whitfield,T.W., Greven,M.C., Pierce,B.G., Dong,X., Kundaje,A., Cheng,Y. *et al.* (2012) Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.*, **22**, 1798–1812.
7. Serandour,A.A., Avner,S., Percevault,F., Demay,F., Bizot,M., Lucchetti-Miganeh,C., Barloy-Hubler,F., Brown,M., Lupien,M., Metivier,R. *et al.* (2011) Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res.*, **21**, 555–565.
8. Gifford,C.A. and Meissner,A. (2012) Epigenetic obstacles encountered by transcription factors: reprogramming against all odds. *Curr. Opin. Genet. Dev.*, **22**, 409–415.
9. Tahiliani,M., Koh,K.P., Shen,Y., Pastor,W.A., Bandukwala,H., Brudno,Y., Agarwal,S., Iyer,L.M., Liu,D.R., Aravind,L. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
10. Ito,S., Shen,L., Dai,Q., Wu,S.C., Collins,L.B., Swenberg,J.A., He,C. and Zhang,Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, **333**, 1300–1303.
11. Maurano,M.T., Wang,H., John,S., Shafer,A., Canfield,T., Lee,K. and Stamatoyannopoulos,J.A. (2015) Role of DNA methylation in modulating transcription factor occupancy. *Cell Rep.*, **12**, 1184–1195.
12. Feldmann,A., Ivanek,R., Murr,R., Gaidatzis,D., Burger,L. and Schubeler,D. (2013) Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS Genet.*, **9**, e1003994.
13. Serandour,A.A., Avner,S., Oger,F., Bizot,M., Percevault,F., Lucchetti-Miganeh,C., Palierne,G., Gheeraert,C., Barloy-Hubler,F., Peron,C.L. *et al.* (2012) Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic Acids Res.*, **40**, 8255–8265.
14. Hon,G.C., Song,C.X., Du,T., Jin,F., Selvaraj,S., Lee,A.Y., Yen,C.A., Ye,Z., Mao,S.Q., Wang,B.A. *et al.* (2014) 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol. Cell*, **56**, 286–297.
15. Robinson,J.L., Holmes,K.A. and Carroll,J.S. (2013) FOXA1 mutations in hormone-dependent cancers. *Front. Oncol.*, **3**, 20.
16. Yu,J., Yu,J., Mani,R.S., Cao,Q., Brenner,C.J., Cao,X., Wang,X., Wu,L., Li,J., Hu,M. *et al.* (2010) An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell*, **17**, 443–454.
17. Yu,J., Cao,Q., Yu,J., Wu,L., Dallol,A., Li,J., Chen,G., Grasso,C., Cao,X., Lonigro,R.J. *et al.* (2010) The neuronal repellent SLIT2 is a target for repression by EZH2 in prostate cancer. *Oncogene*, **29**, 5370–5380.
18. Song,C.X., Clark,T.A., Lu,X.Y., Kislyuk,A., Dai,Q., Turner,S.W., He,C. and Korlach,J. (2012) Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. *Nat. Methods*, **9**, 75–77.
19. Wang,D., Garcia-Bassets,I., Benner,C., Li,W., Su,X., Zhou,Y., Qiu,J., Liu,W., Kaikkonen,M.U., Ohgi,K.A. *et al.* (2011) Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature*, **474**, 390–394.
20. Jin,H.J., Zhao,J.C., Wu,L., Kim,J. and Yu,J. (2014) Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. *Nat. Commun.*, **5**, 3972.
21. Song,C.X., Szulwach,K.E., Fu,Y., Dai,Q., Yi,C., Li,X., Li,Y., Chen,C.H., Zhang,W., Jian,X. *et al.* (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.*, **29**, 68–72.
22. Fong,K.W., Leung,J.W., Li,Y., Wang,W., Feng,L., Ma,W., Liu,D., Songyang,Z. and Chen,J. (2013) MTR120/KIAA1383, a novel microtubule-associated protein, promotes microtubule stability and ensures cytokinesis. *J. Cell Sci.*, **126**, 825–837.
23. Xu,Y., Xu,C., Kato,A., Tempel,W., Abreu,J.G., Bian,C., Hu,Y., Hu,D., Zhao,B., Cerovina,T. *et al.* (2012) Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. *Cell*, **151**, 1200–1213.
24. Jin,C., Lu,Y., Jelinek,J., Liang,S., Estecio,M.R., Barton,M.C. and Issa,J.P. (2014) TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells. *Nucleic Acids Res.*, **42**, 6956–6971.

The Application of Clinical Genetics

Dovepress
open access to scientific and medical research

Open Access Full Text Article

REVIEW

# Influence of oncogenic transcription factors on chromatin conformation and implications in prostate cancer

This article was published in the following Dove Press journal:
The Application of Clinical Genetics
12 May 2014
Number of times this article has been viewed

Yeqing Angela Yang[1]
Jung Kim[1]
Jindan Yu[1,2]

[1]Division of Hematology/Oncology, Department of Medicine, [2]Robert H Lurie Comprehensive Cancer Center, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA

**Abstract:** In recent years, facilitated by rapid technological advances, we are becoming more adept at probing the molecular processes, which take place in the nucleus, that are crucial for the hierarchical regulation and organization of chromatin architecture. With an unprecedented level of resolution, a detailed atlas of chromosomal structures (histone displacement, variants, modifications, chromosome territories, and DNA looping) and mechanisms underlying their establishment, provides invaluable insight into physiological as well as pathological phenomena. In this review, we will focus on prostate cancer, a prevalent malignancy in men worldwide, and for which a curative treatment strategy is yet to be attained. We aim to catalog the most frequently observed oncogenic alterations associated with chromatin conformation, while emphasizing the TMPRSS2-ERG fusion, which is found in more than one-half of prostate cancer patients and its functions in compromising the chromatin landscape in prostate cancer.

**Keywords:** chromatin conformation, ERG, prostate cancer

## Introduction to chromatin organization and three-dimensional topology

The nucleus is a fascinating organelle within a cellular entity, not only due to the fact that it contains the entire genetic blueprint required for a cell to survive and propagate, but – more importantly – how it is capable of organizing this vast sea of information in an efficient and effective manner. It has been known that the human genome consists of more than 3 billion base pairs, and in fact the total deoxyribonucleic acid (DNA) in a diploid human cell would sum up to approximately 2 m in length when completely stretched.[1] Moreover, the extent of compaction for metaphase chromosome is estimated to be between 10,000- and 20,000-fold.[2]

To achieve this high level of proficiency and accuracy, the nucleus employs multiple levels of packaging methods to generate what is known as the higher-order structure of chromatin, which is composed of a combination of DNA and proteins that intertwine together to separate genes into regulatory hubs and to form a three-dimensional (3D) topology best suited for a cell's functions.

About 40 years ago, the use of electron microscopy enabled identification of the classical beads-on-a-string type of structure of DNA, which has been generally accepted as the basic level of chromatin organization.[3–6] Further demonstrated by biochemical and X-ray diffraction studies,[5,7] the chromatin has been described to be formed by repeating units of nucleosomes, octameric structures consisting of four different histone proteins (two each of H2A, H2B, H3, and H4), which are wound by an estimated number of 147 base pairs of DNA, giving rise to 1.7 superhelical turns.[7,8]

Correspondence: Jindan Yu
Division of Hematology/Oncology, Department of Medicine, Robert H Lurie Comprehensive Cancer Center, Northwestern University, Feinberg School of Medicine, 303 East Superior Street, Lurie 5-117, Chicago, IL 60611, USA
Tel +1 312 503 1761
Fax +1 312 503 0189
Email jindan-yu@northwestern.edu

The advent of fluorescence in situ hybridization (FISH) technology has provided evidence for the nonrandom spatial organization of the genome, allowing visualization of position and interaction of chromosomes, chromatin domains, as well as individual genes. It was revealed that gene density is one of the indicators of nuclear positional organization, which is present generally in a radial pattern where gene-dense chromosomal regions prefer to congregate in the nuclear interior, while gene-poor regions are located around the nuclear periphery.[9] It has been observed that chromosomes are segregated into subnuclear compartments, known as chromosome territories,[10] where the edge of the nucleus is host to mainly repressed genes packed into heterochromatin form,[11,12] and the nuclear interior is concentrated in early replicating DNA and frequently transcribed genes.[13,14] In addition, FISH experiments have also demonstrated that during differentiation, specific loci can reposition either toward or away from the nuclear periphery, which is concordant with repression or activation of those nearby genes.[15] More recently, a series of chromosome conformation capture (3C)-based approaches, which can achieve a high-resolution interrogation of the chromatin landscape, further confirmed that intrachromosomal associations in metazoan genome can serve to concentrate and segregate active gene-rich and gene-poor domains.[16]

It is known that – indeed – there are topologically associated domains (TADs) that are pervasive throughout the genome and function to compartmentalize the genome into local and distinct regions, therefore modulating gene expression.[17]

## Role of chromatin organization in gene transcription

The structure of chromatin has been well-known to associate with the status of gene transcription. As early as the 1980s, scientists were able to demonstrate that the mere presence of nucleosomes can inhibit initiation by ribonucleic acid polymerase II (RNAPII) and thus stall transcription.[18] The mechanisms for regulation of the chromatin structure with respect to gene transcription are diverse, and may involve histone displacement, histone variant incorporation, post-translational modifications, chromosome territories, and DNA looping (Figure 1).[19] Each of these mechanisms has its unique influence on chromatin conformation, which in turn dictates gene transcription status.

While packaging of the DNA into nucleosomes can inhibit transcription in vitro, this stereochemical constraint may be relieved by structural changes in nucleosomes.[18] Histones have been observed to exhibit high turnover properties from the core nucleosome. It is reported that histone dimers of H2A and H2B are relatively more susceptible to displacement when compared to H3 and H4.[20] Results from biochemical and genetic studies consistently reinforce the notion that histone eviction from the nucleosome typically occurs at promoters during gene activation, and such process may be mediated by events including but not limited to adenosine triphosphate (ATP)-dependent chromatin remodeling, as well as histone chaperones.[21]

For instance, chromatin remodeling complexes, such as switch/sucrose nonfermentable (SWI/SNF)[22–24] and chromatin structure remodeling (RSC) complex,[25,26] and additionally active RNAP II[27] can all take part in evicting H2A and H2B to assist nucleosome unraveling. Thus, in a stepwise manner, these chromatin remodeling complexes can mediate repositioning[28] or ejection[29] of nucleosomes at promoters to initiate transcription activation. Moreover, histone chaperone proteins (Asf1,[30,31] Nap1,[32] and nucleophosmin[33]), which act by sequestering the evicted histones to prevent their reincorporation into the nucleosome, are also an indispensable component for proper histone displacement and ultimately gene transcription.

In addition to the physical exchange of histone proteins, the incorporation of variant histones can also lead to modifications in chromatin structure and transcriptional regulation. Unlike canonical core histones, generally these unconventional histone proteins are distinguished by the fact that they are expressed outside of the S phase and their deposition into the nucleosome is deemed DNA replication-independent.[34]

As a result of changes in their amino acid sequence, variant forms of histones could acquire divergent biophysical properties predisposing them to localize in specific regions of the genome. One prominent histone variant is H2A.Z, which is an alternative form of H2A, and differs from its counterpart in that its N-terminal tail sequence and several key internal residues, which can effectively alter its ability to interact with H2B as well as the H3/H4 tetramer that eventually manifests in reduced nucleosome stability.[35,36] The deposition of H2A.Z is reportedly carried out by ATP-dependent histone exchange reactions through SWR1,[37] or by the aforementioned chaperone protein Nap1.[38] Another well-studied histone variant is H3.3, and in spite of the fact that it only differs in four amino acids from its canonical form H3, H3.3 has its distinct deposition pattern where it is preferentially enriched in transcriptionally active chromatin and regulatory sites.[39,40]

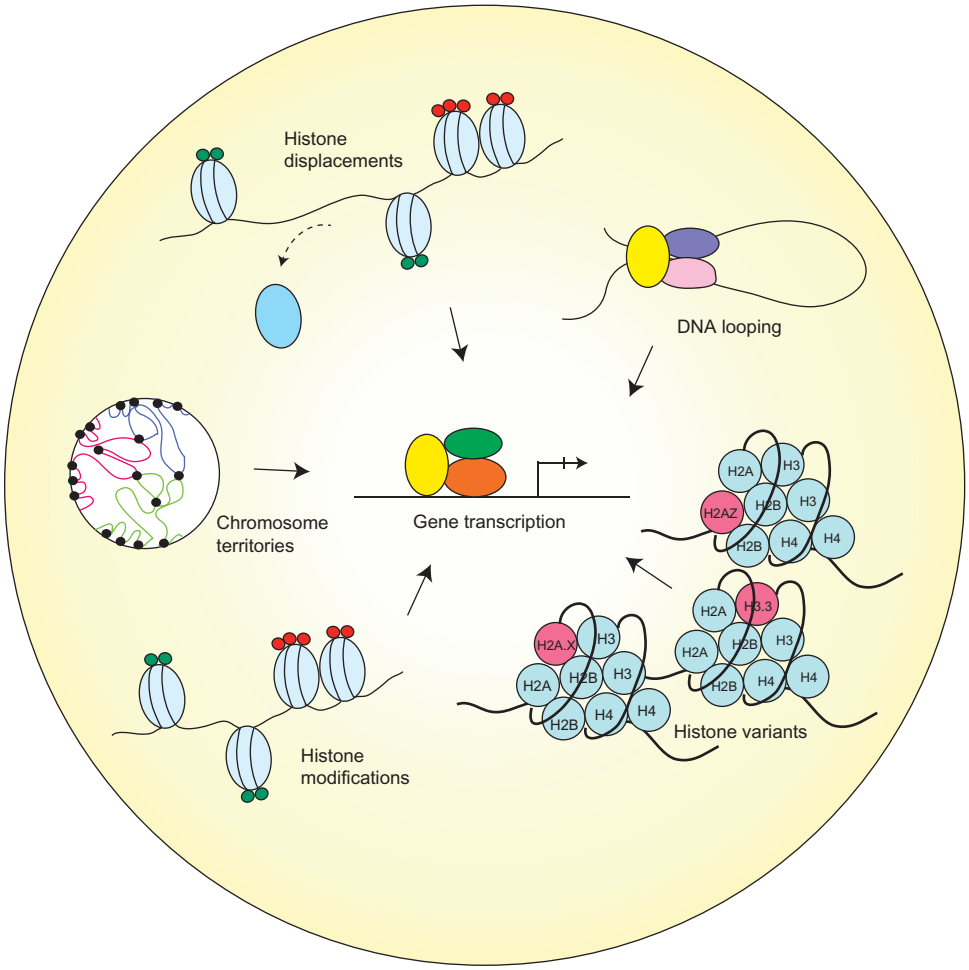



**Figure 1** Different chromatin remodeling regulates gene transcription.
**Note:** Various architectures of the chromatin, histone displacement, DNA looping, histone variants, histone modification, and chromosome territories, regulate gene transcription.

On the other hand, certain variants, such as macroH2A (mH2A), have the ability to repress gene transcription by remodeling the chromatin to impede RNAPII binding. The name mH2A is derived from the structural feature of this histone variant, which contains a large nonhistone region (NHR), known as the macro domain, on its N-terminus.[41] As a consequence, the NHR of mH2A alters nucleosome structure and interferes with the transcription machinery.[42]

Furthermore, a significant category of mechanisms contributing to chromatin organization is posttranslational modifications (PTMs) on histone proteins. There has been extensive research conducted to compile and characterize existing histone modifications, depicting a close relationship between histone PTMs and chromatin structure. Some of the most widely studied histone PTMs include acetylation, methylation, phosphorylation, ubiquitination, and sumoylation. They covalently modify the N-terminal and/or the C-terminal histone tails, while affecting the globular domains at a lesser extent.[43]

These various forms of histone marks generate a code that can be interpreted by specialized proteins to regulate gene expression or to mediate DNA repair.[44] Modifications that reflect in active transcription have been elucidated and include acetylation of H3 and H4, and di- or trimethylation of H3 at lysine position 4 (H3K4me2 or me3). In contrast, modifications that instigate inactivation of transcription include methylation at H3K9 and H3K27.[19]

In eukaryotes, individual chromosomes can occupy spatially defined territories in the interphase nucleus, and repositioning of these genomic regions has an impact on the regulation of gene expression. FISH analysis has shown that chromosome territories adjoin at their borders to create boundaries between chromatin domains. More recently, it is demonstrated that TADs are enclosed by sharp boundaries

enriched for the insulator-binding protein CTCF, as well as the heterochromatin mark H3K9me3.[17] Since boundaries of these topological domains display properties of classical insulator and barrier features, it is therefore suggested that TADs may be linked to transcriptional control.

Concordantly, another study reported that the positions of TADs align with repressive epigenetic marks, as well as lamina-associated domains, and disrupting a TAD boundary can lead to the long-range deregulation in gene expression during X-chromosome inactivation.[45] Therefore, the evidence is convincing that TADs indeed play a role in shaping transcriptional landscapes by clearly defining which sequences belong to the same regulatory network.

Last, as DNA is packaged inside the nucleus, long-range chromatin interactions inevitably occur and – as a result – form loop structures, a majority of which take place between cis-regulatory elements and promoters. It is reported that the dynamic alterations of chromatin looping can either activate or suppress gene expression by facilitating the interactions between enhancers or silencers and their target genes.

One study revealed that only approximately 7% of looping is bridging its nearest gene, reflecting that this chromatin structure is not restrained by genomic proximity and is capable of engaging promoters with distal sites to form complex networks.[46] At the same time, these long-range interactions are not inhibited by CTCF and cohesin occupancy,[46] which argues against previous notions that CTCF's binding to insulator sequences may prevent promoter-enhancer interactions.

Moreover, evidence suggests that the enhancer-promoter loop interactions are formed, in a cell type-specific manner, prior to the binding of transcription factors, indicating their critical role in laying the groundwork for transcriptional control during lineage specification.[47] Furthermore, in terms of thermodynamic properties of DNA looping, it is understood that this mechanism of bringing together multiple components into one functional unit serves to simultaneously increase specificity and affinity and reduce transcriptional noise.[48]

## Role of chromatin conformation in cancer

Due to the crucial role chromatin structure has on determining gene transcription, it is intuitive that chromatin conformation could be manipulated during oncogenic transformation of cancerous cells. It has been demonstrated that under the employment of tumor cells, these chromatin organization machineries become deregulated, disrupting the 3D architecture and undermining the genomic integrity. One of the most recurring phenomena that is associated with cancer development is chromosomal translocations.[49] In the past several decades, a copious number of translocation events have been identified to play pivotal roles in development of a wide range of hematological malignancies as well as solid tumors, which have in turn been utilized as valuable diagnostic and prognostic markers.

Aside from chromosomal translocations, a myriad of events have been implicated in cancer, most of which are deviations from the physiological occurrences of chromatin organization discussed previously. Here, we will catalog the most significant aberrations pertinent to chromatin topology that contribute to cancer development, with a particular emphasis on prostate cancer (Figure 2).

The Philadelphia chromosome is recognized as one the most prominent cancer-associated cytogenetic abnormality that was first reported by Nowell and Hungerford in 1960.[50] It is a highly frequent oncogenic event found in more than 90% of chronic myelogenous leukemia. The translocation is characterized by a reciprocal interchange between chromosome 9 and chromosome 22, which inopportunely generates a BCR-ABL tyrosine kinase gene fusion product.[51] As a result of juxtaposing the breakpoint cluster region (BCR) promoter with the coding region of the *ABL* gene, the hyperactive BCR-ABL fusion protein confers myeloproliferative properties and leads to leukemogenesis.[52] Clinical successes obtained through pharmacological therapies directly inhibiting the activity of BCR-ABL (eg, imatinib mesylate) have provided a promising paradigm in which chromosomal organization could be a critical target for cancer development and, certainly, cancer treatment.

It was not until recently, however, that chromosomal translocations have been identified in solid tumors. In 2005, Tomlins et al made the breakthrough discovery of the fusion of the *TMPRSS2* and *ERG* genes in prostate cancer.[53] According to their study, a striking proportion of 50% of prostate cancers were found to contain a merged product of the 5′ untranslated region of *TMPRSS2* (21q22), an androgen-regulated gene, and the protein-coding sequences of ERG (21q22), an erythroblast transformation-specific (ETS) transcription factor (Figure 2). The *TMPRSS2-ERG* rearrangement has been confirmed to be present in 36%–78% of prostate cancers.[54] In addition, other members of the ETS family, including *ETV1* (7p21), *ETV4* (17q21), and *ETV5* (3q28), were also uncovered as fusion partners with TMPRSS2 in prostate cancer, but they were detected in lower frequency.[55] Unlike the BCR-ABL translocation, the fusion between *TMPRSS2* and *ETS* genes does not generate a chimeric protein, but instead it promotes

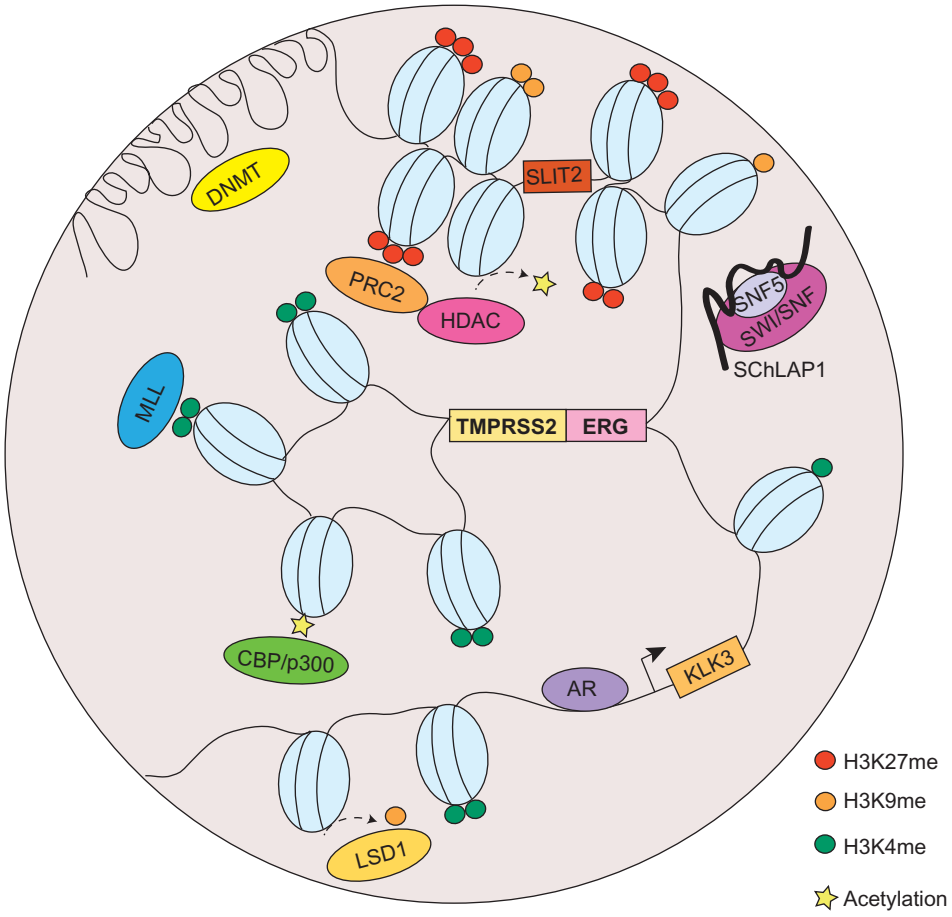



**Figure 2** Chromatin organization aberrations in prostate cancer.
**Note:** Chromatin organizations are altered in prostate cancer through DNA looping, histone PTMs, ncRNAs, and chromosomal translocations, which differentially regulate gene expression.
**Abbreviations:** DNMT, DNA methyltransferase; PTMs, posttranslational modifications; RNA, ribonucleic acid; ncRNA, noncoding RNA; PRC2, polycomb repressive complex 2; HDAC, histone deacetylase; SWI/SNF, switch/sucrose nonfermentable; MLL, mixed-lineage leukemia; AR, androgen receptor; LSD1, lysine-specific demethylase 1.

the overexpression of oncogenic factors directed by a corrupted promoter element. While solely *TMPRSS2* has been identified as a fusion partner of *ERG*, other 5′ partners of *ETS* genes have also been observed. These include androgen-induced genes *SLC45A3, KLK2, CANT1*, and *NDRG1*, and an endogenous retroviral element *HERV-K_22q11.23*, which are functionally comparable to TMPRSS2, as well as androgen-repressed gene *C15orf21*.[56–58]

It was also reported that rearrangements in the rapidly accelerated fibrosarcoma (RAF) pathway also occur in advanced prostate cancer (SLC45A3-BRAF, ESRP1-RAF1), which can be targeted by RAF kinase inhibitors.[59] Moreover, a recent study was able to identify a median of 90 rearrangements in seven prostate cancer tumor samples.[60] Examples of disrupted genes due to rearrangement include *CADM2*, which is a cell adhesion molecule, and phosphatase and tensin homolog (*PTEN*), a well-established tumor suppressor, as well as a PTEN-interacting protein, *MAGI2*. These findings depict a convoluted network of genomic rearrangements and chromatin conformation, which synergistically confer deregulated gene expressions and contribute to tumorigenesis.

In addition to chromosomal translocations, modifications to histone could also place a huge impact on the 3D structure of chromatin and has been widely implicated in cancer. In prostate cancer, H3K4 methylation and H3K27 methylation are among the most extensively investigated histone PTMs; while the former is generally associated with activation of proto-oncogenes, the latter is associated with silencing of tumor suppressors. The repressive epigenetic PTM, H3K27 trimethylation (H3K27me3), has been found to be significantly enriched in promoters of numerous tumor suppressor genes (eg, *ADRB2*,[61] *SLIT2*,[62] *DAB2IP*,[63,64] etc), in metastatic prostate cancer. Meanwhile, H3K9me1 and me2, generally accompanied by heterochromatin assembly,[65] are also implicated in prostate cancer. Demethylation of H3K9 has been reported to reflect in derepression of AR-regulated genes.[66]

H3K4 mono- and dimethylation (H3K4me1, H3K4me2) have been thought of as markers for enhancer sites in directing the androgen receptor (AR) transcriptional program, by facilitating AR binding directly or indirectly through the recruitment of coactivators, such as FOXA1, GATA2, and MED1.[67] Moreover, an endeavor combining high-resolution nucleosome positioning with histone marks mapping showed strong evidence that H3K4me2-containing nucleosomes spaced 250–450 bp (base pair) apart can flank binding sites of AR prior to its ligand-mediated activation, while the binding site is occluded by a well-positioned nucleosome. Following AR activation, nucleosomes with altered H3K4me2 marks become destabilized at AR binding sites and are comparably more stable at the two flanking loci.[68]

In addition, the study revealed that the labile H2A.Z variant was more likely to be present in the central nucleosome relative to the flanking nucleosomes, which further contributes to reduced stability of the nucleosome occupied at the AR binding site. Also, it has been shown that androgen treatment can increase the level of H2A.Z and that the incorporation of H2A.Z in enhancer and proximal promoter sites of the AR-induced gene prostate-specific antigen (*PSA*; or *KLK3*) can poise the gene for activation by AR.[69]

Established and maintained by protein–protein interaction between transcription factors bound at enhancers and at promoters,[70] DNA looping and chromatin compartmentalization are essential processes governing gene transcription; hence, they are a frequent target for disruption during cancer development. In the case of prostate cancer, AR-mediated chromatin looping has been a longtime research interest in the field, and extensive efforts have been devoted to elucidate the process of how AR signaling may lead to changes in chromatin conformation during prostate tumorigenesis. Studies using chromatin immunoprecipitation (ChIP) techniques showed a striking feature of AR genome-wide binding pattern that, approximately 86%–95% of AR localization occurs in nonpromoter regions.[67,71]

This evidence strongly indicates that AR, as a transcription factor, is able to direct its specific transcriptional program from a distance – sometimes, even hundreds of kilobases –away from its target gene. Therefore, it is plausible to presume that a looping model is the mechanism by which AR can regulate its targets from afar. In fact, this model has been proven to be true through 3C-based assays, which demonstrated that distal AR enhancer regions form long-range physical contacts with transcription start sites of AR-regulated genes, such as *PSA* and *TMPRSS2*,[72,73] as well as *UBE2C*, which is a critical enzyme involved in promoting growth of castration-resistant prostate cancer.[67,74]

## Oncogene-mediated alterations in chromatin conformation

Oncogenes have long been implicated in cancer through chromatin alterations, and one route they take is histone modification. It was discovered in 2002 that EZH2, the enzymatic component of the polycomb repressive complex 2 (PRC2), was among the most upregulated genes in prostate cancer.[75] The tumorigenic role of EZH2 has been well-documented, and it involves epigenetic silencing of tumor suppressors and developmental regulators to maintain a dedifferentiated state for cancer cells.[76] EZH2 catalyzes trimethylation of H3K27, creating repressive chromatin structures over long genomic distances.[77,78] It also recruits several other players, such as PRC1, DNA methyltransferase (DNMT), and histone deacetylase (HDAC) (Figure 2), which are concordantly upregulated in prostate cancer.[79–81] It has been revealed that 50% of hypermethylated genes in prostate cancer display preestablished EZH2-mediated H3K27me3 marks, which then leads to de novo DNA methylation.[82] Therefore, EZH2 acts in concert with additional epigenetic enzymes to implement chromatin compaction in a cooperative manner.

Furthermore, methylation at H3K9 has also shown to be deregulated in prostate cancer, through perturbed activities of lysine-specific demethylase 1 (LSD1) (Figure 2).[66] However, the functions of LSD1 in prostate cancer appear to multifaceted, since it is capable to demethylate not only H3K9, but also H3K4. Since H3K4me1 and me2 are essential marks on AR enhancer sites, erasing these modifications consequently result in gene repression.[83] Moreover, an exome-sequencing study recently revealed that several members of the mixed-lineage leukemia (MLL) complex (MLL, MLL2, and ASH2L), which acts as H3K4-specific methyltransferase, can physically interact with AR and are significantly mutated in prostate cancer[84] (Figure 2). From a translational standpoint, pharmacological targeting of these histone-modifying enzymes has been envisaged and shown clinical triumph.[85]

While histone modifications are carried out by specific enzymes, the molecular process underlying the formation of chromatin looping may be effected through a network of coregulators (eg, MED12, SRC-1, p300/CBP, BRG1, etc) that are collectively responsible for sustaining the loop structure.[1] Additionally, the GATA, OCT, PAX, NKX, and LEF family proteins have been observed to have sequence motifs near nuclear hormone receptors, including AR and estrogen

receptor (ER).[86] Disruption of chromosomal structures, therefore, can significantly impair proper gene transcription. A central protein in AR/ER signaling, the pioneer factor forkhead box A1 (FOXA1), has been regarded as a key mediator of AR/ER transcription regulation through chromatin remodeling and recruitment of AR/ER to target sites.[87]

The fact that FOXA1 is overexpressed and mutated in hormone-dependent cancers, prostate cancer, and breast cancer, is in concordance with its predominant role in directing AR/ER signaling to drive cancer development.[88,89] In addition, knowledge about the multiprotein Mediator complex, which is well-known for its role in bridging enhancer and promoter into close proximity,[90] has also contributed to our understanding about chromosome looping involving AR, wherein the silencing of a Mediator subunit MED1 can significantly impair AR transactivation.[91]

Moreover, in the past decade, the AR signaling pathway has also been shown to play essential roles in altering chromatin conformation, primarily due to its involvement in a majority of chromosomal translocations identified in prostate cancer. Through FISH analyses, it was discovered that androgen stimulation can induce the spatial proximity between *TMPRSS2* and *ERG*, thus highly augmenting the probability of forming a fusion product when under the stress of DNA double-strand breaks.[92] Further evidence demonstrated that AR binding at specific intronic loci near break sites in *TMPRSS2*, *ERG*, and *ETV1* could result in rapid formation of intra- and interchromosomal interactions that in turn generate enough spatial proximity to predispose the genes for translocation.

In addition, the ensuing modifications of chromatin architecture sensitize these regions to genotoxic stress, making the translocation loci particularly susceptible to double-stranded breaks. The liganded AR, upon binding to DNA, can recruit enzymes – including activation-induced cytidine deaminase (AID) and LINE-1 repeat-encoded ORF2 endonuclease, as well as topoisomerase II beta (TOP2B) to create double-stranded breaks at break sites, which then become ligated through nonhomologous end joining.[93,94]

Aside from proteins playing an oncogenic function, there has been emerging evidence that long noncoding RNAs (lncRNAs; >200 nt)[95] may also adversely affect chromatin structures. For instance, it is recently reported that *HOTAIR*, a 2.2 kb lncRNA residing in the *HOXC* locus, serves as a crucial interface between DNA and the chromatin-modifying complex PRC2. As a result, in breast cancer, an overexpression of *HOTAIR* is causally linked to alterations in the chromatin state reimposed by PRC2 occupancy, consequently permitting a gene expression program that is conducive to cell motility and invasion by silencing key metastasis suppressor genes.[96]

Another prominent member of the lncRNAs that is recently implicated in cancer is *SChLAP1*, which was identified as an overexpressed gene in prostate tumor samples.[97] Similar to *HOTAIR, SChLAP1*, in context of prostate cancer, can promote cancer invasion and metastasis. The molecular mechanisms underlying *SChLAP1's* oncogenic function were also connected to a chromatin modifying complex, namely SWI/SNF (Figure 2). Through antagonizing the genomic binding of SWI/SNF, *SChLAP1* significantly impairs the transcriptional program directed by SWI/SNF that, as alluded to previously, is a complex that utilizes ATP to mobilize nucleosomes and remodel chromatin.[97]

Numerous links have been established between SWI/SNF and carcinogenesis;[98] however, the latest discovery of *SChLAP1*, in addition to *HOTAIR*, sheds new light on the mechanistic basis of how deregulation of lncRNAs may result in defective chromatin organization, which ultimately contributes to oncogenesis. Furthermore, a recent study reports that two lncRNAs overexpressed in prostate cancer, *PCGEM1* and *PRNCR1* (*PCAT8*), can bind with AR and facilitate the enhancer-promoter loop formation required for AR transcriptional regulation. It was demonstrated that these lncRNAs can promote AR activation in a hormone-independent environment, providing novel mechanistic insight into the pathogenesis of castration-resistant prostate cancer.[99]

Therefore, taken together, these lines of encouraging evidence keep propelling scientists forward to continuously uncover novel mechanisms associated with deregulation of chromosomal organization – to provide beneficial insight into strategies for diagnosing as well as treating cancer.

## ERG overexpression and chromatin conformation

ERG is overexpressed in prostate cancer due to AR-mediated changes in chromosome rearrangement. As a consequence, ERG overexpression, in turn, can also lead to chromatin structure alterations, which further contribute to prostate cancer development in a feed-forward vicious cycle. The function of ERG involves physical interaction with a number of cofactors as well as transcription factors, including AR – ultimately leading to a transcriptional program favoring the dedifferentiation, invasion, and neoplastic transformation of prostate epithelial cells.[100]

To characterize the molecular crosstalk between ERG and AR, studies showed that ERG can effectively attenuate AR signaling by the direct transcriptional repression of AR, and

additionally ERG occupancy at AR target genes correlates with negative regulation, which is potentiated by ERG-induced EZH2 activity.[76,101] It is becoming increasingly clear that ERG complexes with other molecules to coordinately organize chromatin structure. Several interacting partners of ERG have been identified in recent years, including EZH2[71,102] and HDAC1,[103] which – along with ERG – form a repressor coregulatory network that is important for mediating androgen response in prostate cancer.[101] This notion is further supported by the fact that ERG overexpression also dictates changes in the genomewide DNA methylation landscape,[104] reflecting a complex regulatory program directed by ERG to impose structural alterations in the overall 3D chromatin topology.

Employing a combination of advanced technologies, including Hi-C, ChIP-seq, and RNA-seq, and integrative bioinformatic analyses, Rickman et al[100,105] showed that an overexpression of ERG could induce dramatic changes in 3D chromatin topology, corresponding to the changes in the expression of a group of genes implicated in aggressive prostate cancer. Since ERG binding strongly associated with hotspots of differential chromatin interactions, an upregulation of ERG when fused to TMPRSS2 upon androgen stimulation consequently leads to altered regulation of transcription events.

Among these ERG-regulated genes are genes associated with invasion and migration (eg, *FYN, PLAU, MMP3, MMP9, LEF1*, and *miR200c*) and urogenital development (eg, *HOXA, B, C* gene cluster members, *PYGO1*, and *NKX3.1*).[100,106–108]

## Concluding remarks and therapeutic implications

Over the past several decades, we have witnessed a plethora of pioneering studies that established the essential role of chromatin conformation during normal biological processes and oncogenic cellular transformations.

Through investigations of molecular mechanisms governing the alterations in chromatin architecture, researchers have been able to strategically design therapeutic agents which, by abolishing the enzymatic activity of certain chromatin-modifying proteins, to achieve the correct 3D chromatin topology. Several drugs were recently approved by the FDA due to their improved efficacy in prolonging survival and reduced toxicity compared to conventional chemotherapy. Some prominent examples include DNA methylation inhibitor azacitidine (Vidaza®) and decitabine (Dacogen®) and HDAC inhibitors vorinostat (Zolinza®) and romidepsin (Istodax®), which were FDA-approved successively in the last 10 years, for the treatment of myelodysplastic syndrome and cutaneous T-cell lymphoma, respectively.[109]

Currently, clinical trials are being conducted to examine the pharmacological efficacy of DNMT and HDAC inhibitors in prostate cancer, as adjuvant therapies to complement androgen deprivation.[110] In addition, a wide range of chemical inhibitors targeting enzymes, such as EZH2 (eg, DZNep[111] and GSK126[112]) and LSD1 (eg, TCP[113] and ORY-1001[85]), have demonstrated promising potential in various in vitro and in vivo studies for multiple cancer types.[85]

These molecules hold hopeful prospective for treatment of prostate cancer, in which oncogenic contributors to chromosomal abnormalities are abundant. It is anticipated that future pharmaceutical therapies aimed to restore the physiological activity level of key chromatin modulators would provide desirable curative effects.

## Acknowledgments

This work was supported in part by: the National Institutes of Health/National Cancer Institute training grant T32CA09560 (to YAY); National Institutes of Health R01CA172384 (to JY); the US Army Medical Research and Materiel Command grant W81XWH-13-1-0319 (to JY); and the Research Scholar Award RSG-12-085-01 (to JY) from the American Cancer Society.

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Wu D, Zhang C, Shen Y, Nephew KP, Wang Q. Androgen receptor-driven chromatin looping in prostate cancer. *Trends Endocrinol Metab*. 2011;22(12):474–480.
2. Woodcock CL, Ghosh RP. Chromatin higher-order structure and dynamics. *Cold Spring Harb Perspect Biol*. 2010;2(5):a000596.
3. Oudet P, Gross-Bellard M, Chambon P. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell*. 1975;4(4):281–300.
4. Hewish DR, Burgoyne LA. Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem Biophys Res Commun*. 1973;52(2):504–510.
5. Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. *Science*. 1974;184(4139):868–871.
6. Olins AL, Olins DE. Spheroid chromatin units (v bodies). *Science*. 1974;183(4122):330–332.
7. Finch JT, Lutter LC, Rhodes D, et al. Structure of nucleosome core particles of chromatin. *Nature*. 1977;269(5623):29–36.
8. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*. 1997;389(6648):251–260.
9. Bickmore WA. The spatial organization of the human genome. *Annu Rev Genomics Hum Genet*. 2013;14:67–84.

10. Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*. 2001;2(4): 292–301.
11. Croft JA, Bridger JM, Boyle S, Perry P, Teague P, Bickmore WA. Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol*. 1999;145(6):1119–1131.
12. Bridger JM, Bickmore WA. Putting the genome on the map. *Trends Genet*. 1998;14(10):403–409.
13. Xing Y, Johnson CV, Moen PT Jr, McNeil JA, Lawrence J. Nonrandom gene organization: structural arrangements of specific pre-mRNA transcription and splicing with SC-35 domains. *J Cell Biol*. 1995; 131(6 Pt 2):1635–1647.
14. Carter KC, Bowman D, Carrington W, et al. A three-dimensional view of precursor messenger RNA metabolism within the mammalian nucleus. *Science*. 1993;259(5099):1330–1335.
15. Kosak ST, Skok JA, Medina KL, et al. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science*. 2002;296(5565):158–162.
16. Simonis M, Klous P, Splinter E, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet*. 2006;38(11):1348–1354.
17. Dixon JR, Selvaraj S, Yue F, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485(7398):376–380.
18. Knezetic JA, Luse DS. The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. *Cell*. 1986;45(1): 95–104.
19. Li B, Carey M, Workman JL. The role of chromatin during transcription. *Cell*. 2007;128(4):707–719.
20. Kimura H, Cook PR. Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *J Cell Biol*. 2001;153(7):1341–1353.
21. Workman JL. Nucleosome displacement in transcription. *Genes Dev*. 2006;20(15):2009–2017.
22. Owen-Hughes T, Utley RT, Côté J, Peterson CL, Workman JL. Persistent site-specific remodeling of a nucleosome array by transient action of the SWI/SNF complex. *Science*. 1996;273(5274):513–516.
23. Bruno M, Flaus A, Stockdale C, Rencurel C, Ferreira H, Owen-Hughes T. Histone H2A/H2B dimer exchange by ATP-dependent chromatin remodeling activities. *Mol Cell*. 2003;12(6):1599–1606.
24. Phelan ML, Schnitzler GR, Kingston RE. Octamer transfer and creation of stably remodeled nucleosomes by human SWI-SNF and its isolated ATPases. *Mol Cell Biol*. 2000;20(17):6380–6389.
25. Lorch Y, Zhang M, Kornberg RD. RSC unravels the nucleosome. *Mol Cell*. 2001;7(1):89–95.
26. Lorch Y, Maier-Davis B, Kornberg RD. Chromatin remodeling by nucleosome disassembly in vitro. *Proc Natl Acad Sci U S A*. 2006;103(9):3090–3093.
27. Kireeva ML, Walter W, Tchernajenko V, Bondarenko V, Kashlev M, Studitsky VM. Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription. *Mol Cell*. 2002;9(3):541–552.
28. Whitehouse I, Flaus A, Cairns BR, White MF, Workman JL, Owen-Hughes T. Nucleosome mobilization catalysed by the yeast SWI/SNF complex. *Nature*. 1999;400(6746):784–787.
29. Lorch Y, Zhang M, Kornberg RD. Histone octamer transfer by a chromatin-remodeling complex. *Cell*. 1999;96(3):389–392.
30. Adkins MW, Howar SR, Tyler JK. Chromatin disassembly mediated by the histone chaperone Asf1 is essential for transcriptional activation of the yeast PHO5 and PHO8 genes. *Mol Cell*. 2004;14(5):657–666.
31. Schwabish MA, Struhl K. Asf1 mediates histone eviction and deposition during elongation by RNA polymerase II. *Mol Cell*. 2006;22(3):415–422.
32. Walter PP, Owen-Hughes TA, Côté J, Workman JL. Stimulation of transcription factor binding and histone displacement by nucleosome assembly protein 1 and nucleoplasmin requires disruption of the histone octamer. *Mol Cell Biol*. 1995;15(11):6178–6187.
33. Swaminathan V, Kishore AH, Febitha KK, Kundu TK. Human histone chaperone nucleophosmin enhances acetylation-dependent chromatin transcription. *Mol Cell Biol*. 2005;25(17):7534–7545.
34. Kamakaka RT, Biggins S. Histone variants: deviants? *Genes Dev*. 2005;19(3):295–310.
35. Zhang H, Roberts DN, Cairns BR. Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell*. 2005;123(2):219–231.
36. Raisner RM, Hartley PD, Meneghini MD, et al. Histone variant H2A.Z marks the 5′ ends of both active and inactive genes in euchromatin. *Cell*. 2005;123(2):233–248.
37. Mizuguchi G, Shen X, Landry J, Wu WH, Sen S, Wu C. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science*. 2004;303(5656):343–348.
38. Park YJ, Chodaparambil JV, Bao Y, McBryant SJ, Luger K. Nucleosome assembly protein 1 exchanges histone H2A-H2B dimers and assists nucleosome sliding. *J Biol Chem*. 2005;280(3):1817–1825.
39. Henikoff S. Labile H3.3+H2A.Z nucleosomes mark 'nucleosome-free regions'. *Nat Genet*. 2009;41(8):865–866.
40. Yuen BT, Knoepfler PS. Histone H3.3 mutations: a variant path to cancer. *Cancer Cell*. 2013;24(5):567–574.
41. Pehrson JR, Fried VA. MacroH2A, a core histone containing a large nonhistone region. *Science*. 1992;257(5075):1398–1400.
42. Doyen CM, An W, Angelov D, et al. Mechanism of polymerase II transcription repression by the histone variant macroH2A. *Mol Cell Biol*. 2006;26(3):1156–1164.
43. Berger SL. The complex language of chromatin regulation during transcription. *Nature*. 2007;447(7143):407–412.
44. Kouzarides T. Chromatin modifications and their function. *Cell*. 2007;128(4):693–705.
45. Nora EP, Lajoie BR, Schulz EG, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*. 2012;485(7398): 381–385.
46. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature*. 2012;489(7414): 109–113.
47. Jin F, Li Y, Dixon JR, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*. 2013;503(7475):290–294.
48. Vilar JM, Saiz L. DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr Opin Genet Dev*. 2005;15(2):136–144.
49. Göndör A. Dynamic chromatin loops bridge health and disease in the nuclear landscape. *Semin Cancer Biol*. 2013;23(2):90–98.
50. Nowell PC, Hungerford DA. Chromosome studies on normal and leukemic human leukocytes. *J Natl Cancer Inst*. 1960;25:85–109.
51. Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature*. 1973;243(5405):290–293.
52. Lugo TG, Pendergast AM, Muller AJ, Witte ON. Tyrosine kinase activity and transformation potency of bcr-abl oncogene products. *Science*. 1990;247(4946):1079–1082.
53. Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science*. 2005;310(5748):644–648.
54. Demichelis F, Fall K, Perner S, et al. TMPRSS2:ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. *Oncogene*. 2007;26(31):4596–4599.
55. Clark JP, Cooper CS. ETS gene fusions in prostate cancer. *Nat Rev Urol*. 2009;6(8):429–439.
56. Tomlins SA, Laxman B, Dhanasekaran SM, et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature*. 2007;448(7153):595–599.
57. Hermans KG, Bressers AA, van der Korput HA, Dits NF, Jenster G, Trapman J. Two unique novel prostate-specific and androgen-regulated fusion partners of ETV4 in prostate cancer. *Cancer Res*. 2008;68(9): 3094–3098.

58. Pflueger D, Rickman DS, Sboner A, et al. N-myc downstream regulated gene 1 (NDRG1) is fused to ERG in prostate cancer. *Neoplasia*. 2009;11(8):804–811.
59. Palanisamy N, Ateeq B, Kalyana-Sundaram S, et al. Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. *Nat Med*. 2010;16(7):793–798.
60. Berger MF, Lawrence MS, Demichelis F, et al. The genomic complexity of primary human prostate cancer. *Nature*. 2011;470(7333):214–220.
61. Yu J, Cao Q, Mehra R, et al. Integrative genomics analysis reveals silencing of beta-adrenergic signaling by polycomb in prostate cancer. *Cancer Cell*. 2007;12(5):419–431.
62. Yu J, Cao Q, Yu J, et al. The neuronal repellent SLIT2 is a target for repression by EZH2 in prostate cancer. *Oncogene*. 2010;29(39):5370–5380.
63. Chen H, Tu SW, Hsieh JT. Down-regulation of human DAB2IP gene expression mediated by polycomb Ezh2 complex and histone deacetylase in prostate cancer. *J Biol Chem*. 2005;280(23):22437–22444.
64. Min J, Zaslavsky A, Fedele G, et al. An oncogene-tumor suppressor cascade drives metastatic prostate cancer by coordinately activating Ras and nuclear factor-kappaB. *Nat Med*. 2010;16(3):286–294.
65. Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science*. 2001;292(5514):110–113.
66. Metzger E, Wissmann M, Yin N, et al. LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature*. 2005;437(7057):436–439.
67. Wang Q, Li W, Zhang Y, et al. Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell*. 2009;138(2):245–256.
68. He HH, Meyer CA, Shin H, et al. Nucleosome dynamics define transcriptional enhancers. *Nature Genet*. 2010;42(4):343–347.
69. Dryhurst D, McMullen B, Fazli L, Rennie PS, Ausió J. Histone H2A.Z prepares the prostate specific antigen (PSA) gene for androgen receptor-mediated transcription and is upregulated in a model of prostate cancer progression. *Cancer Lett*. 2012;315(1):38–47.
70. Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. *Cell*. 2011;144(3):327–339.
71. Yu J, Yu J, Mani RS, et al. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell*. 2010;17(5):443–454.
72. Wang Q, Carroll JS, Brown M. Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Mol Cell*. 2005;19(5):631–642.
73. Wang Q, Li W, Liu XS, et al. A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. *Mol Cell*. 2007;27(3):380–392.
74. Chen Z, Zhang C, Wu D, et al. Phospho-MED1-enhanced UBE2C locus looping drives castration-resistant prostate cancer growth. *EMBO J*. 2011;30(12):2405–2419.
75. Varambally S, Dhanasekaran SM, Zhou M, et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature*. 2002;419(6907):624–629.
76. Yu J, Yu J, Rhodes DR, et al. A polycomb repression signature in metastatic prostate cancer predicts cancer outcome. *Cancer Res*. 2007;67(22):10657–10663.
77. Francis NJ, Kingston RE. Mechanisms of transcriptional memory. *Nat Rev Mol Cell Biol*. 2001;2(6):409–421.
78. Zhao JC, Yu J, Runkle C, et al. Cooperation between Polycomb and androgen receptor during oncogenic transformation. *Genome Res*. 2012;22(2):322–331.
79. Cao Q, Mani RS, Ateeq B, et al. Coordinated regulation of polycomb group complexes through microRNAs in cancer. *Cancer Cell*. 2011;20(2):187–199.
80. Weichert W, Röske A, Gekeler V, et al. Histone deacetylases 1, 2 and 3 are highly expressed in prostate cancer and HDAC2 expression is associated with shorter PSA relapse time after radical prostatectomy. *Br J Cancer*. 2008;98(3):604–610.
81. Nelson WG, Yegnasubramanian S, Agoston AT, et al. Abnormal DNA methylation, epigenetics, and prostate cancer. *Front Biosci*. 2007;12:4254–4266.
82. Schlesinger Y, Straussman R, Keshet I, et al. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet*. 2007;39(2):232–236.
83. Cai C, He HH, Chen S, et al. Androgen receptor gene expression in prostate cancer is directly suppressed by the androgen receptor through recruitment of lysine-specific demethylase 1. *Cancer Cell*. 2011;20(4):457–471.
84. Grasso CS, Wu YM, Robinson DR, et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature*. 2012;487(7406):239–243.
85. Helin K, Dhanak D. Chromatin proteins and modifications as drug targets. *Nature*. 2013;502(7472):480–488.
86. Lupien M, Brown M. Cistromics of hormone-dependent cancer. *Endocr Relat Cancer*. 2009;16(2):381–389.
87. Lupien M, Eeckhoute J, Meyer CA, et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell*. 2008;132(6):958–970.
88. Robinson JL, Holmes KA, Carroll JS. FOXA1 mutations in hormone-dependent cancers. *Front Oncol*. 2013;3:20.
89. Jin HJ, Zhao JC, Ogden I, Bergan RC, Yu J. Androgen receptor-independent function of FoxA1 in prostate cancer metastasis. *Cancer Res*. 2013;73(12):3725–3736.
90. Kagey MH, Newman JJ, Bilodeau S, et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*. 2010;467(7314):430–435.
91. Wang Q, Sharma D, Ren Y, Fondell JD. A coregulatory role for the TRAP-mediator complex in androgen receptor-mediated gene expression. *J Biol Chem*. 2002;277(45):42852–42858.
92. Mani RS, Tomlins SA, Callahan K, et al. Induced chromosomal proximity and gene fusions in prostate cancer. *Science*. 2009;326(5957):1230.
93. Lin C, Yang L, Tanasa B, et al. Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell*. 2009;139(6):1069–1083.
94. Haffner MC, Aryee MJ, Toubaji A, et al. Androgen-induced TOP2B-mediated double-strand breaks and prostate cancer gene rearrangements. *Nat Genet*. 2010;42(8):668–675.
95. Djebali S, Davis CA, Merkel A, et al. Landscape of transcription in human cells. *Nature*. 2012;489(7414):101–108.
96. Gupta RA, Shah N, Wang KC, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*. 2010;464(7291):1071–1076.
97. Prensner JR, Iyer MK, Sahu A, et al. The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat Genet*. 2013;45(11):1392–1398.
98. Roberts CW, Orkin SH. The SWI/SNF complex – chromatin and cancer. *Nat Rev Cancer*. 2004;4(2):133–142.
99. Yang L, Lin C, Jin C, et al. lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature*. 2013;500(7464):598–602.
100. Rickman DS, Soong TD, Moss B, et al. Oncogene-mediated alterations in chromatin conformation. *Proc Natl Acad Sci U S A*. 2012;109(23):9083–9088.
101. Chng KR, Chang CW, Tan SK, et al. A transcriptional repressor co-regulatory network governing androgen response in prostate cancers. *EMBO J*. 2012;31(12):2810–2823.
102. Kunderfranco P, Mello-Grand M, Cangemi R, et al. ETS transcription factors control transcription of EZH2 and epigenetic silencing of the tumor suppressor gene Nkx3.1 in prostate cancer. *PLoS One*. 2010;5(5):e10547.
103. Iljin K, Wolf M, Edgren H, et al. TMPRSS2 fusions with oncogenic ETS factors in prostate cancer involve unbalanced genomic rearrangements and are associated with HDAC1 and epigenetic reprogramming. *Cancer Res*. 2006;66(21):10242–10246.

104. Kron K, Trudel D, Pethe V, et al. Altered DNA methylation landscapes of polycomb-repressed loci are associated with prostate cancer progression and ERG oncogene expression in prostate cancer. *Clin Cancer Res*. 2013;19(13):3450–3461.
105. Elemento O, Rubin MA, Rickman DS. Oncogenic transcription factors as master regulators of chromatin topology: a new role for ERG in prostate cancer. *Cell Cycle*. 2012;11(18):3380–3383.
106. Tomlins SA, Laxman B, Varambally S, et al. Role of the TMPRSS2-ERG gene fusion in prostate cancer. *Neoplasia*. 2008;10(2):177–188.
107. Kim J, Wu L, Zhao JC, Jin HJ, Yu J. TMPRSS2-ERG gene fusions induce prostate tumorigenesis by modulating microRNA miR-200c. *Oncogene*. 2013.
108. Wu L, Zhao JC, Kim J, Jin HJ, Wang CY, Yu J. ERG is a critical regulator of Wnt/LEF1 signaling in prostate cancer. *Cancer Res*. 2013;73(19):6068–6079.
109. Boumber Y, Issa JP. Epigenetics in cancer: what's the future? *Oncology (Williston Park)*. 2011;25(3):220–226, 228.
110. Lin J, Wang C, Kelly WK. Targeting epigenetics for the treatment of prostate cancer: recent progress and future directions. *Semin Oncol*. 2013;40(3):393–401.
111. Tan J, Yang X, Zhuang L, et al. Pharmacologic disruption of Polycomb-repressive complex 2-mediated gene repression selectively induces apoptosis in cancer cells. *Genes Dev*. 2007;21(9):1050–1063.
112. McCabe MT, Ott HM, Ganji G, et al. EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature*. 2012;492(7427):108–112.
113. Schenk T, Chen WC, Göllner S, et al. Inhibition of the LSD1 (KDM1A) demethylase reactivates the all-trans-retinoic acid differentiation pathway in acute myeloid leukemia. *Nat Med*. 2012;18(4):605–611.

nature COMMUNICATIONS

ARTICLE

Received 26 Oct 2013 | Accepted 25 Apr 2014 | Published 30 May 2014

DOI: 10.1038/ncomms4972

# Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program

Hong-Jian Jin[1], Jonathan C. Zhao[1], Longtao Wu[1], Jung Kim[1] & Jindan Yu[1,2]

The pioneering factor FOXA1 opens chromatin to facilitate androgen receptor (AR) binding to prostate-specific genes. How FOXA1 controls the AR cistrome, however, is incompletely understood. Here we show that AR directly binds chromatin through the androgen response elements (AREs). FOXA1 is not required for AR–chromatin interaction, but instrumental in recruiting AR to low-affinity half-AREs by opening local chromatin around adjacent FKHD sites. Too much FOXA1 creates excessive open chromatin regions, which serve as reservoirs that retain AR via abundant half-AREs, thereby reducing its availability for specific sites. FOXA1 downregulation, by contrast, relinquishes AR to permissively bind AREs across the genome, resulting in substantial AR-binding events and AR target gene expression even in the absence of androgen. Taken together, our data illustrate the mechanistic details by which cooperativity and equilibrium with FOXA1 define AR cistrome and reveal a previously unknown function of FOXA1 in inhibiting AR signalling and castration-resistant prostate cancer growth.

[1] Division of Hematology/Oncology, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois 60611, USA. [2] Robert H. Lurie Comprehensive Cancer Center, Northwestern University Feinberg School of Medicine, Chicago, Illinois 60611, USA. Correspondence and requests for materials should be addressed to J.Y. (email: jindan-yu@northwestern.edu).

The androgen receptor (AR), a hormone-activated transcription factor of the nuclear receptor family, is a key regulator of prostatic gene expression[1,2] and plays pivotal roles in prostate differentiation and function[3]. Aberrant elevation of AR signalling, on the other hand, is a critical driver of malignant transformation of the prostate gland and as such androgen-deprivation therapy has been a mainstay treatment of prostate cancer (PCa)[4]. Moreover, AR continues to be upregulated in advanced PCa and its expression and activity remain required for the growth of castration-resistant PCa (CRPC) in an androgen-depleted environment[5]. Studies have shown that AR can become transactivated in CRPC by a number of different mechanisms including AR amplification, AR mutation and alterations in cofactor proteins[6].

AR is a DNA-binding protein that, upon androgen stimulation, binds cis-regulatory elements that harbour the androgen response element (ARE). In addition, like other hormonal receptors such as estrogen receptor, AR often binds distal enhancers, rather than promoters, and regulates target genes through chromatin looping[7]. In order to enable efficient targeting of the AR pathway, numerous studies have attempted to determine the downstream molecules of AR transcriptional regulation. Coupling chromatin immunoprecipitation with next-generation sequencing (ChIP-seq), we and others have revealed tens of thousands of AR-binding events across the human genome at high resolution[8,9]. Motif analyses confirmed ARE as the most prevalent DNA motif within AR-binding sites (ARBS), supporting its essential role in mediating AR–DNA interaction. Previous studies have shown that AR–DNA-binding profile is also tightly regulated by an extensive list of cofactors, one key member of which is FOXA1.

FOXA1, also known as HNF-3, is a forkhead family transcriptional factor. Possessing winged-helix DNA-binding domains similar to linker histones, FOXA1 is able to access compact chromatin to form high-affinity DNA binding with the FKHD motif and subsequently acts to open up the local nucleosomal domain[10]. FOXA1 is highly expressed in prostate epithelial cells and is critical for the regulation of prostate ductal morphogenesis and epithelial cell differentiation and maturation[11]. This function is largely mediated by the ability of FOXA1 to tightly control AR-modulated transcriptional regulation of prostatic genes[12]. FOXA1 was shown to induce the expression of AR target genes such as PSA by co-occupying the FKHD motif that locates immediately adjacent to the AR-bound ARE motif within the PSA enhancer. Gao *et al.*[12] have also shown that the FOXA1 protein physically interacts with the AR protein, thereby acting as an AR-collaborating cofactor. Specifically, the DNA-binding domain/hinge region of AR directly interacts with the forkhead domain of FOXA1 (refs 12,13).

Genome-wide location analyses comparing FOXA1-bound genomic regions in prostate and breast cancer cells have revealed cell type-specific recruitment, which subsequently dictates distinct AR and estrogen receptor chromatin-binding patterns in prostate and breast cells, respectively[14]. This model is strongly supported by the findings that FOXA1 binds DNA even in the absence of androgen and preoccupies a majority of the ARBS stimulated by androgen[14–16]. This is especially appealing as FOXA1 is known to act as a pioneer factor that opens up compact chromatin to facilitate the recruitment of other transcription factors including hormonal receptors[10,17–19]. Cumulatively, FOXA1 is thus known as a pioneer factor for AR. However, FOXA1 binds substantially more genomic regions than AR and that a majority of FOXA1-binding sites (FXBS) are not co-occupied by AR, arguing that FOXA1 is not sufficient to recruit AR[14–16]. In addition, genome-wide analysis of AR-binding profiles showed that FOXA1 knockdown (KD) resulted in a drastic shift, rather than loss, of ARBS, suggesting that FOXA1 is also not required for AR–chromatin interaction[15,16]. How FOXA1 regulates AR cistrome is not fully understood. Here we present evidence that FOXA1 overall inhibits AR signalling, which is mediated by two mechanisms: (1) FOXA1 reduces AR availability to specific ARE sites by facilitating AR binding to numerous half ARE (hARE) sites located within the proximity of FOXA1-bound FKHD sites, and (2) FOXA1 directly inhibits AR gene expression. We further show that FOXA1-KD results in AR binding, target gene expression and PCa cell growth in the absence of androgen.

## Results

**AR DNA-binding ability is required for chromatin occupancy.** To determine how FOXA1 defines AR-binding profile, we first conducted FOXA1 KD in LNCaP PCa cells and then performed AR and FOXA1 ChIP-Seq in both control and KD cells (Supplementary Fig. 1A). As expected, FXBS were drastically reduced following FOXA1 KD (Fig. 1a). In addition, ChIP-Seq-read intensity of the remaining FXBS was also significantly decreased (Supplementary Fig. 1B). By contrast, AR binding was substantially shifted from some loci to others, confirming reprogramming (Fig. 1b). In addition, we observed an overall increase of total ARBS; while ~30% of the initial ARBS were lost twice as many new ones were gained following the silencing of FOXA1. We categorized all ARBS into three categories: present only in control KD cells (shCtrl-only), only in FOXA1-KD cells (shFOXA1-only) or shared in both. Bioinformatics analysis revealed that the shared ARBS have the strongest ChIP-Seq-read intensity, which was further enhanced upon FOXA1 KD, suggesting that they were initially inhibited by FOXA1 (Fig. 1c,d, Supplementary Fig. 1C). Surprisingly, the shCtrl-only ARBS were in general quite weak, indicating that FOXA1-dependent AR binding is likely mediated by low-affinity chromatin interactions. By contrast, shFOXA1-only ARBS were of much stronger binding intensity, thus revealing a FOXA1-independent, high-affinity, AR–DNA interaction mechanism. These global patterns were further confirmed through ChIP–qPCR analysis of a representative set of loci from each category (Supplementary Fig. 1D) and also in independent replicate experiments (Supplementary Fig. 2A–D).

Our data thus strongly suggest a shifted and gained AR programme in PCa cells following the silencing of FOXA1. To examine whether this may be a phenomenon specific to one cell line, we performed FOXA1 KD in LAPC4 cells using the same small hairpin RNA (shRNA) targeting FOXA1 (Supplementary Fig. 2E,F). ChIP–qPCR analysis confirmed that FOXA1 KD resulted in drastic increase of AR binding to the chromatin. This finding was further validated using an independent small interfering RNA (siRNA) that transiently silence FOXA1 in LAPC4 cells, thereby excluding any nonspecific off-target effects (Supplementary Fig. 2G,H). Similarly, we found AR overexpression could also lead to AR recruitment to the gained ARBS, supporting that they are authentic AR targets (Supplementary Fig. 2I,J). In addition, FOXA1 overexpression strongly suppressed these AR-binding events, further supporting its inhibitory role (Supplementary Fig. 2K–M). In addition, FOXA1 was also insufficient in recruiting AR to the DNA since a majority of FXBS were not co-occupied by AR (Supplementary Fig. 3A,B). Collectively, our data showed that FOXA1 is neither required nor sufficient for AR–chromatin binding. Next, we examined what modulates the strong AR–chromatin interaction observed in shared and shFOXA1-only ARBS.

Motif analysis revealed that strong ARBS (shared and shFOXA1-only) were characterized by high (~30%) incidences of full-ARE motif, while only 7% of shCtrl-only ARBS contained an ARE (Fig. 1e, Supplementary Fig. 4A). By contrast, ~45% of

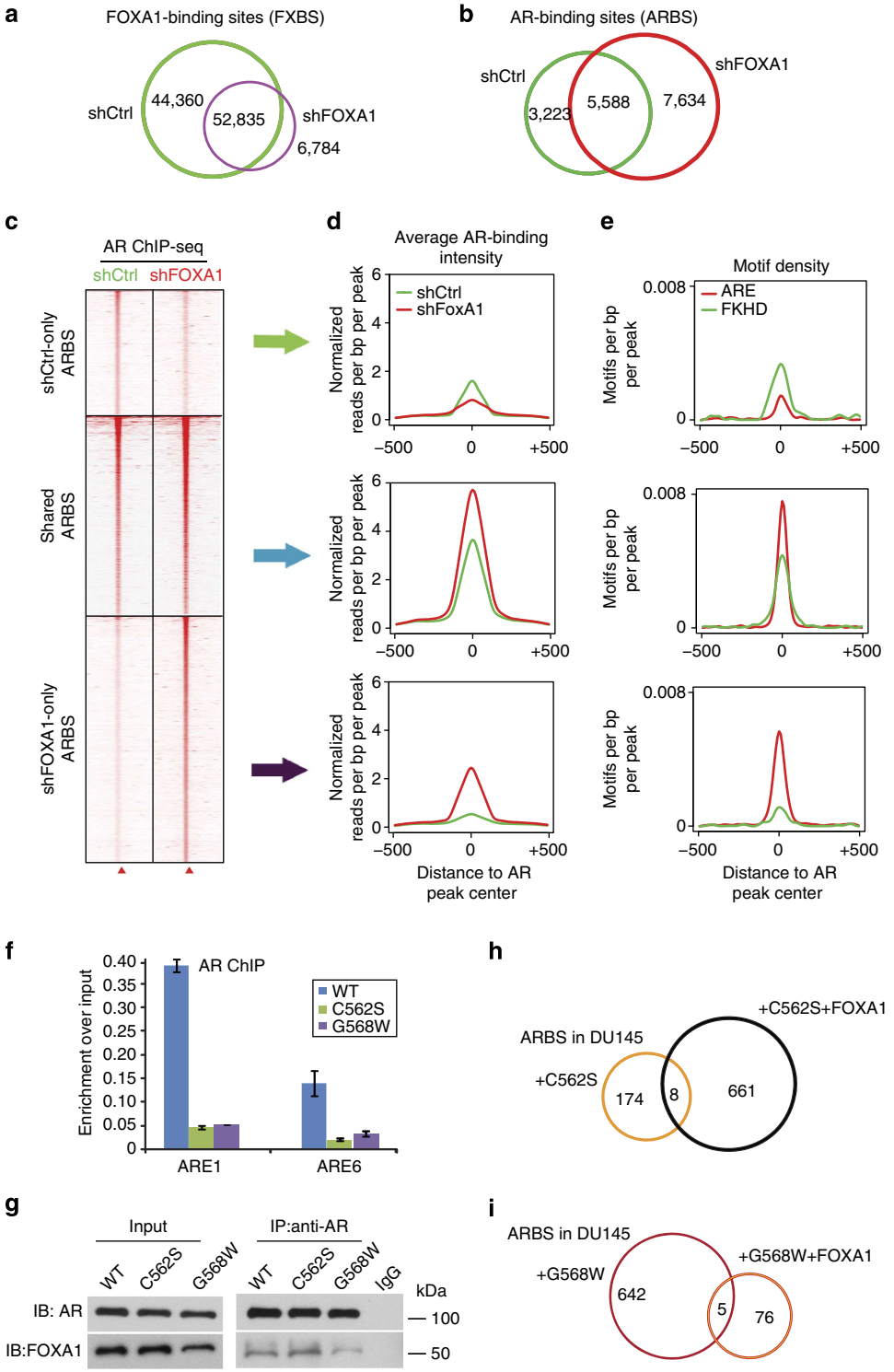



**Figure 1 | The DNA-binding ability of AR is required for its occupancy on the chromatin.** (**a**) Overlap of FXBS. ChIP-Seq of AR and FOXA1 were performed in control (shCtrl) and FOXA1-KD (shFOXA1) LNCaP cells. (**b**) Overlap of ARBS between shCtrl and shFOXA1 LNCaP cells. (**c**) Heatmap view of AR and FOXA1 ChIP-Seq-read intensity around ARBS (±2 kb) detected in shCtrl and shFOXA1 LNCaP cells. ARBS were separated into shCtrl-only, shared and shFOXA1-only. (**d**) Average AR ChIP-seq intensity around ARBS (±500 bp) shown in **c**. (**e**) ARE and FKHD motif density around ARBS (±500 bp) shown in **c**. Motif density was determined by HOMER and normalized to that in a control sequence of equal length. (**f**) AR mutants (C562S and G568W) had substantially lower ability to bind ARE than wt AR. AR ChIP was done in DU145 cells with stable expression of wt or mutant AR. Error bars indicate $n=3$, mean ± s.e.m., $P<0.05$ by two-tailed Student's $t$-test. (**g**) AR mutants (C562S and G568W) remained capable of interacting with FOXA1. Co-immunoprecipitation was carried out in DU145 cells with stable expression of wt or mutant AR infected with FOXA1-expressing adenovirus for 48 h. (**h**,**i**) FOXA1 failed to increase DNA-binding events of mutant AR. Venn diagram shows overlap of ARBS between control- and FOXA1-expressing DU145 + AR (C562S or W568W) cells.

shCtrl-only ARBS harboured a FKHD motif, much more than the 15% found in other categories, being consistent with their FOXA1-dependent binding. Shared ARBS, not surprisingly, contained high incidence (~30 and 55%, respectively) of both ARE and FKHD motifs, which explains their resistance to FOXA1 KD. Moreover, *de novo* motif discovery identified a composite motif containing hARE and FKHD, specifically in shCtrl-only ARBS, suggesting that these weak binding events may be mediated by AR binding to hARE sites, which is facilitated by FOXA1-accessing FKHD motif (Supplementary Fig. 4B).

To test the importance of direct AR–chromatin interaction, we generated AR mutants C562S and G568W that are deficient in DNA binding but remained capable of interacting with the FOXA1 protein (Fig. 1f,g). To assess their ability of DNA binding globally, we infected the AR-negative and FOXA1-negative/low DU145 cells with wild-type (wt) and mutant AR to generate stable clones (DU145 + AR, DU145 + C562S and DU145 + G568W). ChIP-Seq analysis detected very few AR-binding events in the mutant-expressing cells, while ~20,000 AR-binding events were detected in cells expressing wt AR (Supplementary Fig. 5A). Next, we concomitantly overexpressed FOXA1 in these cells (Supplementary Fig. 5B). Remarkably, ChIP-Seq analysis showed that expression of FOXA1, although led to extensive FOXA1 binding to the chromatin as expected, failed to recruit DNA binding-deficient AR mutants to the chromatin (Fig. 1h,i and Supplementary Fig. 5C). Therefore, the DNA-binding ability of AR is absolutely required for its occupancy on the chromatin, supporting its direct interaction with hAREs and full AREs, respectively, in shCtrl-only or shFOXA1-only ARBS.

**FOXA1 reduces AR–chromatin binding**. To engineer an independent model to further examine how FOXA1 regulates AR-binding profile, we overexpressed AR in the FOXA1-negative DU145 cells to generate stable cell lines (DU145 + AR) (Supplementary Fig. 6A). ChIP-Seq analysis revealed abundant AR-binding events in these cells, further supporting that FOXA1 is not required for AR to bind the chromatin (Fig. 2a). We next generated DU145 + AR + FOXA1 cells that express both FOXA1 and AR. Being consistent with the results observed in LNCaP cells, FOXA1 overexpression shifted AR from its original binding sites to the new genomic loci that were now bound by FOXA1 (Supplementary Fig. 6B,C), although the specific AR and FOXA1 programs in DU145 cells differed substantially from those in LNCaP cells (Supplementary Fig. 7). However, these newly acquired FOXA1-dependent ARBS (DU145 + AR + FOXA1-only) were of remarkably lower binding intensity than those present in DU145 + AR cells (shared and DU145 + AR-only ARBS) (Fig. 2b). Moreover, the binding intensity of the shared ARBS was also decreased, further supporting an overall inhibition of AR–chromatin occupancy by FOXA1 overexpression. Being consistent with our findings in LNCaP cells, motif analysis showed that strong ARBS were mediated by ARE motifs, whereas the weaker FOXA1-dependent ARBS (DU145 + AR + FOXA1-only) harboured remarkably more FKHD motifs (Fig. 2c). Overall, FOXA1 overexpression drastically shifted AR binding from high-affinity ARE-containing sites to the genomic loci that harbour FKHD motifs (Supplementary Fig. 8A,B). Next, we wondered whether excessive amount of FOXA1 might further recruit AR away from high-affinity sites by inducing more FXBS.

We engineered DU145 + AR cells to express very high amount of FOXA1 (Supplementary Fig. 6A). ChIP-Seq analysis confirmed that these cells (DU145 + AR + FOXA1$^{high}$) had gained an additional 63,000 FXBS compared with the regular DU145 + AR + FOXA1 cells (Supplementary Fig. 8C). Surprisingly, there were markedly less AR-binding events in these cells; only 4,400 ARBS were detected compared with the 20,000 ARBS found in DU145 + AR or + AR + FOXA1 cells (Fig. 2d). With increased FOXA1 expression, nearly 19,000 ARBS were lost with a minimal gain of 1,000 new ones, demonstrating a remarkable effect of FOXA1 in inhibiting AR-binding events. Moreover, even for the conserved ARBS, the read intensity was also substantially reduced (Fig. 2e and Supplementary Fig. 8D). Intriguingly, these remaining ARBS were not only co-occupied by FOXA1 but also enriched for strong ARE motifs (Supplementary Fig. 8E,F). To validate these important findings, we selected a representative set of genomic loci that contained only ARE, FKHD or both. ChIP–qPCR demonstrated that AR binding to the ARE-only sites was very strong, but drastically reduced following FOXA1 overexpression, and further diminished upon excessive amount of FOXA1 treatment, although these sites were never occupied by FOXA1 (Fig. 2f). Similar inhibition, albeit to a lesser extent, was also observed in ARE + FKHD sites, which were co-occupied by FOXA1 as expected (Fig. 2g). By contrast, AR binding to FKHD-only sites was in general much weaker, increased upon FOXA1 overexpression, but declined following excessive FOXA1 expression (Fig. 2h). Overall, when FOXA1 was overly abundant (in DU145 + AR + FOXA1$^{high}$ cells), AR ChIP-enrichment was very low across all three categories of target regions (Fig. 2f,h). Therefore, medium level of FOXA1 shifted AR from high-affinity ARE sites to facilitate their weak binding to FOXA1-bound FKHD sites, while high FOXA1 attenuated all AR-binding events. Next, we attempted to decipher the mechanisms underlying this inhibition.

**FOXA1 absorbs AR by opening up excessive chromatin regions**. As FOXA1 has been previously shown to open up compact chromatin to facilitate the access of other transcription factors including AR[14], we carried out formaldehyde-assisted isolation of regulatory elements (FAIRE)-seq experiments to assay open chromatin regions. Bioinformatics analysis demonstrated that FOXA1 overexpression not only led to a shift but also a great increase in the number of total open chromatin regions (Fig. 3a). In addition, a substantial proportion of these open chromatin regions were occupied by FOXA1, confirming the role of FOXA1 as a chromatin-opening factor (Supplementary Fig. 9A). To determine how this change in open chromatin relates to the alteration in AR-binding profile, we examined FAIRE-seq signal around ARBS and found that the gained ARBS (DU145 + AR + FOXA1-only) indeed became much more accessible following FOXA1 overexpression (Fig. 3b and Supplementary Fig. 9B). By contrast, the DU145 + AR-only ARBS had drastically reduced chromatin accessibility, which may be due to the loss of AR binding. Interestingly, the FAIRE-seq signal around the shared ARBS remained unaltered, indicating that other mechanisms, such as decreased availability of AR, may be accountable for the attenuation of AR binding around these sites. To validate these global observations, we carried out FAIRE–qPCR and confirmed that, while ARE-only loci became less accessible following FOXA1 overexpression, FKHD-only loci showed dramatically increased accessibility that became several magnitudes higher than the ARE-only sites (Fig. 3c–e). The ARE + FKHD sites, on the other hand, had the highest accessibility, which remained largely unchanged following FOXA1 overexpression.

Next, we investigated how excessive FOXA1 results in immense loss of total AR-binding events as illustrated in Fig. 2d,e. FAIRE-seq data showed that high FOXA1 overexpression opened up an additional 10,000 chromatin sites (Fig. 3f). This, along with the marvellous gain of FXBS (~63,000), would drastically increase the opportunity for low-affinity AR binding, which was in direct

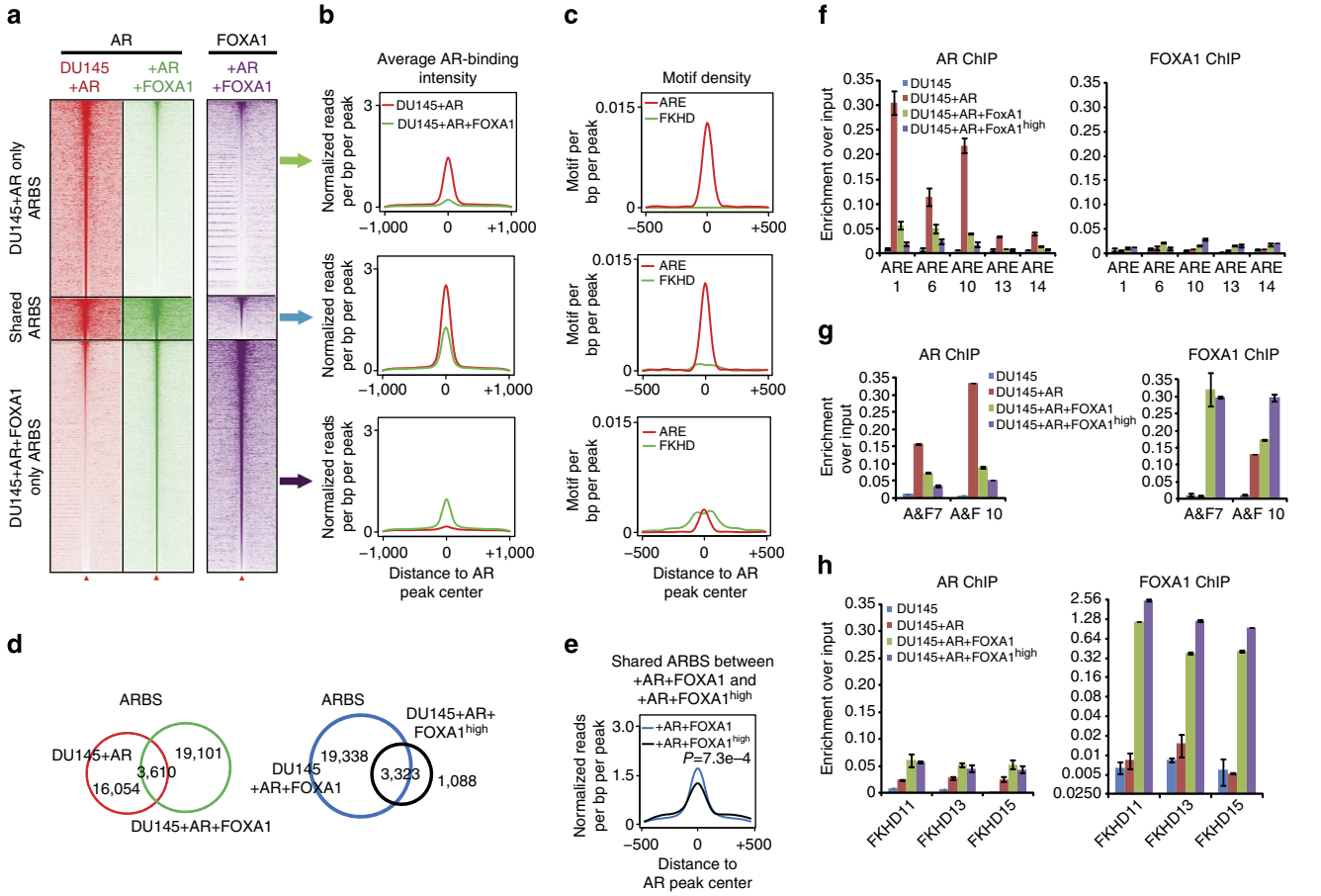



**Figure 2 | FOXA1 reduces AR-chromatin binding.** (**a**) Heatmap view of AR and FOXA1 ChIP-Seq-read intensity around ARBS (±2 kb) that were detected in DU145 + AR and DU145 + AR + FOXA1 stable cells. ARBS were separated into DU145 + AR-only, shared and DU145 + AR + FOXA1-only. (**b**) Average AR ChIP-seq-read intensity around ARBS (±500 bp) shown in **a**. (**c**) ARE and FKHD motif density around ARBS (±500 bp) shown in **a**. Motif density was determined by HOMER and normalized to a control sequence of equal length. (**d**) Balanced expression of FOXA1 resulted in global redistribution of ARBS, whereas excessive FOXA1 led to remarkable inhibition of ARBS. AR ChIP-Seq was carried out in DU145 + AR, DU145 + AR + FOXA1 and DU145 + AR + FOXA1$^{high}$ cells. The cells were generated by infecting DU145 + AR stable cells with LacZ or FOXA1-expressing adenovirus (at 1:2,000 or 1:500 dilution) for 48 h. (**e**) Average intensity of ARBS shared between DU145 + AR + FOXA1 and DU145 + AR + FOXA1$^{high}$ cells. (**f**–**h**) ChIP–qPCR analysis for AR and FOXA1 occupancy at target regions containing ARE-only (**f**), ARE and FKHD (A and F) (**g**) or FKHD-only (**h**) motifs. ChIP was carried out in DU145 + AR, DU145 + AR + FOXA1, and DU145 + AR + FOXA1$^{high}$ cells. Error bars indicate $n = 3$, mean ± s.e.m., $P < 0.05$ by two-tailed Student's *t*-test.

contrast with the observed loss of ~80% of the ARBS. In addition, FAIRE-seq signal around the lost ARBS were only decreased slightly, probably due to increased FOXA1 binding to these sites compensating AR loss (Fig. 3g). This eliminates chromatin compacting as a main cause of the loss of ARBS and suggests alternative mechanisms. To determine whether the lost AR still remains on the chromatin, we performed chromatin fractionation followed by immunoblotting, which demonstrated that, in spite of the drastically decreased AR-binding events detected by ChIP-seq, the amount of chromatin-bound AR was surprisingly not reduced by FOXA1 over-expression (Supplementary Fig. 9C). These suggest that AR proteins remained on the chromatin but their binding was not detectable by ChIP-Seq. This is plausible as we have previously shown that FXBS-mediated AR-binding events were generally very weak (Fig. 2a), which may be further diminished when excessive FOXA1 enables AR to access extensive amount of open chromatin regions. Therefore, FOXA1 opens up chromatin to absorb AR to numerous FXBS via weak interactions, thereby diluting AR and reducing its binding on individual sites.

**FOXA1 interaction enhances AR binding to ARE+FKHD sites.** We next asked what mediates the low-affinity AR–chromatin interaction within the FXBS. To address this, we compared the small portion (~10%) of FXBS that were co-occupied by AR with the majority (90%) that were not (Supplementary Fig. 10A). FOXA1 ChIP-seq data showed that AR tends to co-occupy relatively strong FXBS, probably due to higher accessibility to these regions (Supplementary Fig. 10B). Strikingly, the FXBS that were co-occupied by AR contained much higher occurrence rate of ARE half-sites (Supplementary Fig. 10C,D). By contrast, the occurrence of FKHD motif was similar between FXBS with or without AR co-occupancy. Since hARE has substantially less affinity with AR than full ARE, this explains our earlier observation that AR binding to FXBS is generally weaker. Therefore, FOXA1 binding to FKHD sites increases accessibility and facilitates AR binding to the hARE motifs within these sites. As FOXA1 protein has been shown to directly interact with the AR protein[12], we next asked whether this interaction is required for AR recruitment to the FXBS.

We generated an AR mutant (Q640X) that only expresses the AR protein up to position 640, which has been previously shown

unable to interact with the FOXA1 protein[12,15]. We first confirmed that this mutant was not able to interact with FOXA1 but remained capable of binding ARE-containing DNA (Fig. 4a and Supplementary Fig. 10E). In addition, ChIP-Seq analysis of DU145 cells with stable expression of this mutant (DU145 + AR_Q640X) demonstrated that it retained the ability to bind the chromatin and a majority of its binding sites overlapped with that of wt AR (Fig. 4b). Moreover, FOXA1 overexpression also led to a remarkable shift of the binding profile of the mutant AR, similar to wt AR (Fig. 4c and Supplementary Fig. 10F). Most importantly, this shift was also characterized by a dramatic loss of ARE-mediated high-affinity AR-binding events and a substantial gain of FKHD-mediated, FOXA1-dependent weaker ARBS (Fig. 4d–f). Together, out data suggest that physical interaction with AR protein is not required for FOXA1 to redistribute AR from ARE to FXBS. These results support that, instead of actively recruiting AR through direct interaction, FOXA1 passively enables AR to access the chromatin.

Next, we examined whether this interaction, although not necessary for AR recruitment, might enhance or stabilize AR binding. We first compared the intensity of endogenous ARBS in LNCaP cells that contain ARE-only, FKHD-only, or both (Fig. 4g). Our data confirmed that ARE-mediated binding sites were much stronger than FKHD-mediated ones. In addition, ARBS containing both ARE and FKHD were the strongest, suggesting that FOXA1 co-occupancy in these regions might enhance/stabilize AR binding. Indeed, in FOXA1-KD LNCaP cells, there was no appreciable difference in AR-binding intensity between ARE-only and ARE + FKHD sites, while intensity of both remained superior to that of FKHD-only sites (Fig. 4h).

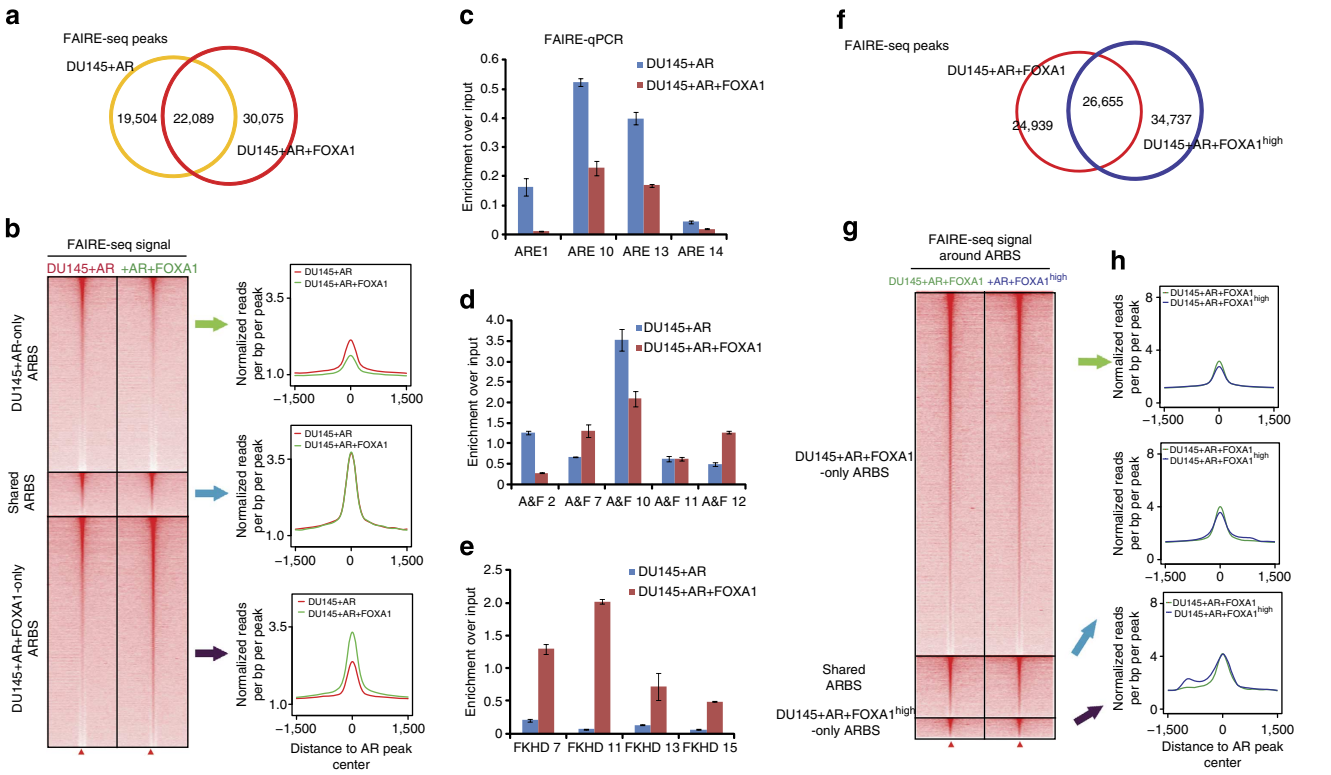
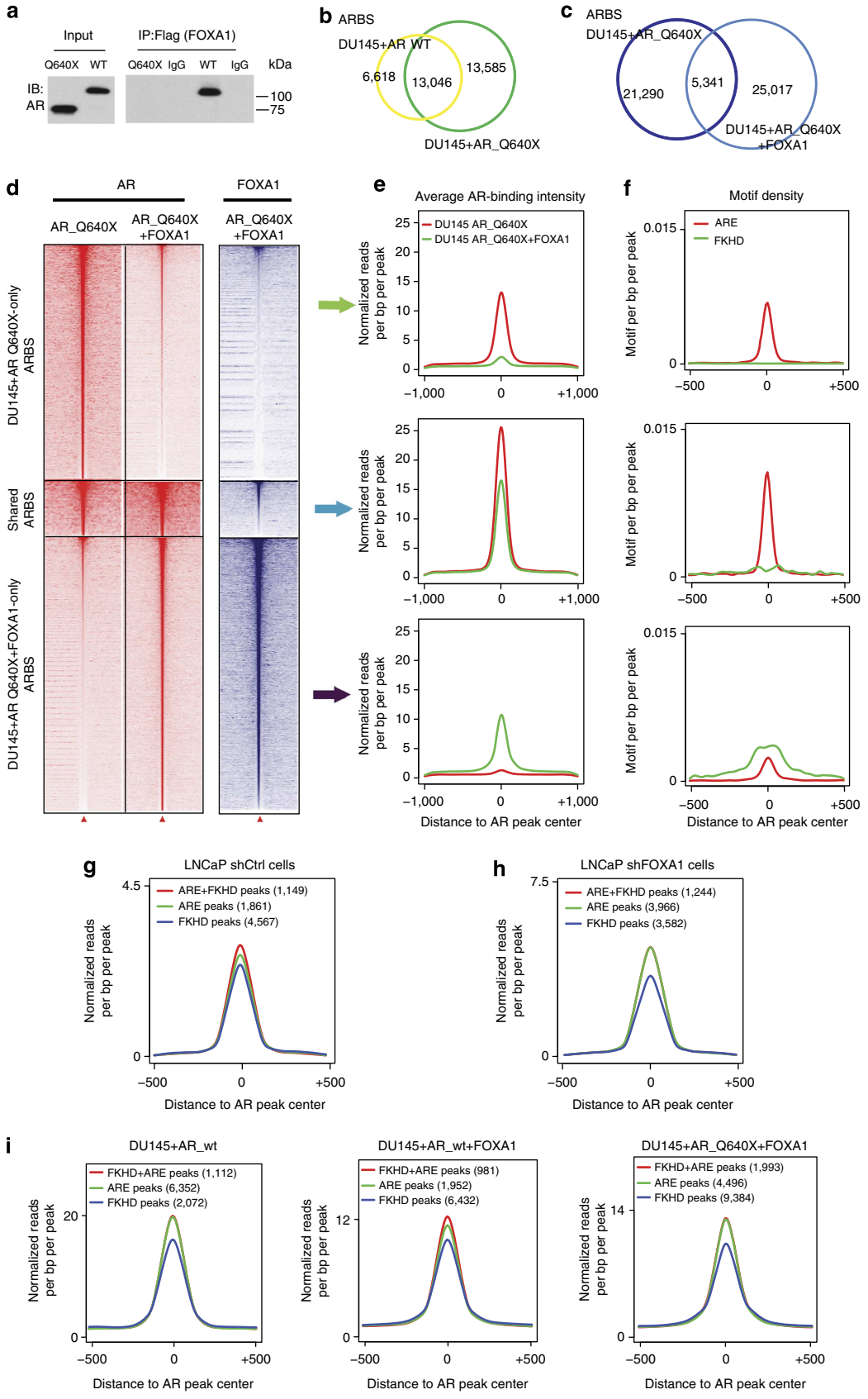



**Figure 3 | FOXA1 opens up chromatin to absorb AR via low-affinity interactions.** (**a**) Overlap between FAIRE-seq peaks identified in DU145 + AR and DU145 + AR + FOXA1 cells. (**b**) Heatmap view of FAIRE-seq read intensity around ARBS (5 kb) detected in DU145 + AR and/or DU145 + AR + FOXA1 stable cells. Shown on the right are average FAIRE-seq counts (±1.5 kb) corresponding to each category of ARBS. (**c**–**e**) FAIRE-qPCR analysis of chromatin accessibility at randomly selected loci from ARBS containing ARE-only (**c**), ARE and FKHD (**d**) or FKHD-only (**e**). FAIRE experiment was carried out in DU145 + AR and DU145 + AR + FOXA1 cells. Overall, $P<0.05$ for ARE-only sites (**c**) and FKHD-only sites (**e**) and $P$ was not significant on average for all A and F sites (**d**). Error bars indicate $n=3$, mean ± s.e.m. $P$-values were calculated by two-tailed Student's $t$-test. (**f**) Overlap between FAIRE-seq peaks detected in DU145 + AR + FOXA1 and in DU145 + AR + FOXA1$^{high}$ cells. (**g**) Heatmap view of FAIRE-seq read intensity around ARBS (5 kb) detected in DU145 + AR + FOXA1 and/or DU145 + AR + FOXA1$^{high}$ cells. Shown on the right is average FAIRE-seq signal (±1.5 kb) corresponding to each category of ARBS.

**Figure 4 | FOXA1 and AR protein interaction mediates preferential AR binding to ARE + FKHD sites.** (**a**) Truncated AR (Q640X) failed to interaction with FOXA1. Co-immunoprecipitation experiments were carried out in 293T cells transfected with wt or Q640X-mutant AR along with FOXA1 for 48 h. (**b**) Overlap of ARBS between DU145 + AR (wt) and DU145 + AR_Q640X cells. (**c**) Ectopic expression of FOXA1 resulted in global redistribution of Q640X-mutant AR binding. Venn diagram shows overlap between ARBS detected in DU145 + AR_Q640X and in DU145 + AR_Q640X + FOXA1 cells. (**d**) Heatmap view of AR and FOXA1 ChIP-Seq-read intensity around ARBS (±2 kb) detected in DU145 + AR_Q640X and in DU145 + AR_Q640X + FOXA1 cells. (**e**) Average AR ChIP-seq intensity around ARBS (±1 kb) shown in **d**. (**f**) ARE and FKHD motif density around ARBS (±500 bp) shown in **d**. (**g**,**h**) Average AR ChIP-seq-read intensity around ARBS detected in LNCaP shCtrl cells (**g**) or in LNCaP shFOXA1 cells (**h**). Based on the motifs that they contain, ARBS were categorized into ARE + FKHD peaks, ARE-only peaks and FKHD-only peaks. (**i**) Average AR ChIP-seq-read intensity around ARBS detected in DU145 + AR_wt, DU145 + AR_wt + FOXA1 or DU145 + AR_Q640X + FOXA1 cells. Based on the motifs that they contain, ARBS were separated into ARE + FKHD peaks, ARE-only peaks and FKHD-only peaks.




To further validate this, we examined ectopically introduced ARBS in DU145 cells. Concordantly, we found that in the absence of FOXA1, the intensity of ARE-only ARBS was the same as those bound to ARE + FKHD sites, both of which were significantly stronger than the FKHD-only sites (Fig. 4i). Interestingly, following FOXA1 expression, ARE + FKHD ARBS again became apparently stronger than ARE-only ARBS, which remained superior to the FKHD-only ones. Most importantly, in

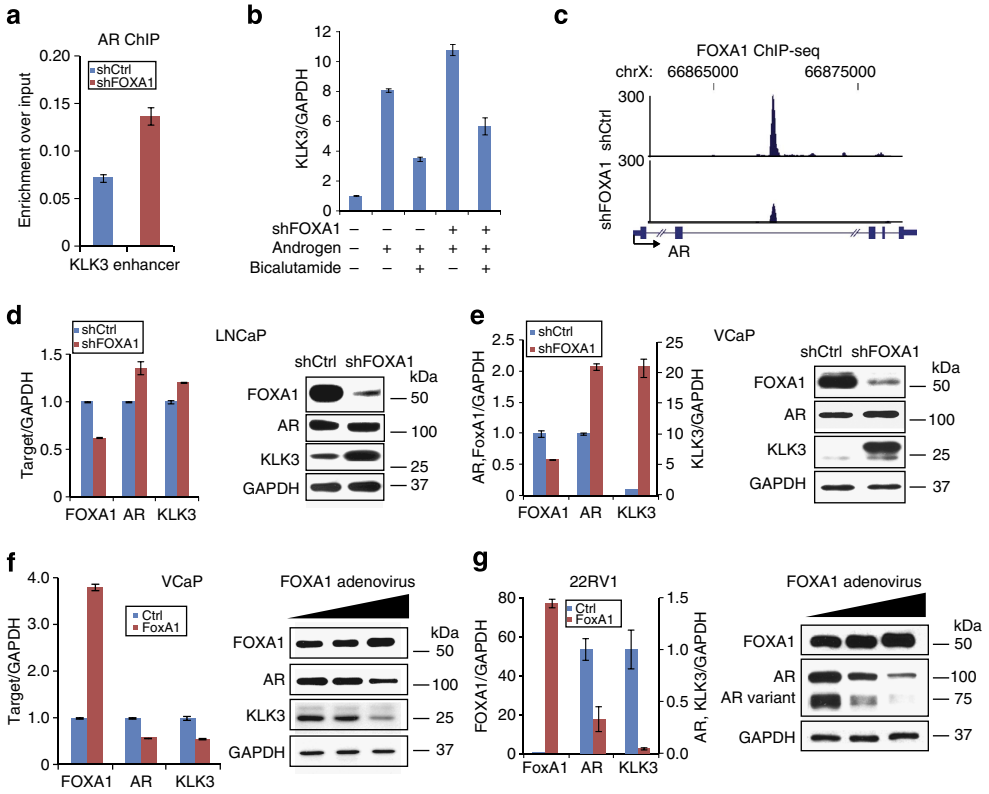



**Figure 5 | FOXA1 inhibits AR gene expression in PCa.** (**a**) AR ChIP was performed in control and shFOXA1 LNCaP cells followed by qPCR analysis of KLK3 enhancer. (**b**) QRT-PCR analysis of KLK3 expression in control and FOXA1-KD LNCaP cells treated with androgen, anti-androgen bicaluatamide or both. (**c**) FOXA1 binding to AR intragenic region in LNCaP cell revealed by ChIP-seq. FOXA1 binding is dramatically decreased by FOXA1 KD. (**d**,**e**) QRT-PCR and immunoblot analysis of FOXA1, AR and KLK3 in control and FOXA1-KD LNCaP (**d**) and VCaP cells (**e**). (**f**,**g**) QRT-PCR and immunoblot analysis of FOXA1, AR and KLK3 in VCaP (**f**) and 22RV1 cells (**g**) following FOXA1 overexpression. For QRT-PCR, cells were infected with adenovirus (at 1:2,000 dilutions) expressing LacZ control or FOXA1 for 48 h. For immunoblot analysis, cells were infected with increasing amount of FOXA1 adenovirus at 1:8,000, 1:2,000 or 1:500 dilutions. Error bars indicate $n=3$, mean ± s.e.m., $P<0.05$ by two-tailed Student's *t*-test.

DU145 + AR_Q640X cells wherein FOXA1 was not able to interact with ARE, the preferential AR binding to ARE + FKHD sites was lost. Therefore, physical interaction between FOXA1 and AR enhances AR binding on FKHD + ARE sites, plausibly by locking in both proteins on the chromatin.

**FOXA1 inhibits AR expression in PCa cells**. In view of the great increase of ARBS following the silencing of FOXA1 in the endogenous LNCaP system, we asked whether AR might be a direct target of FOXA1-mediated transcriptional repression[20]. We first carried out ChIP–qPCR analysis of the prototype AR target gene KLK3, also called PSA. Our data showed that AR binding to the KLK3 enhancer was indeed drastically increased upon FOXA1 KD (Fig. 5a). QRT–PCR analysis further showed that while androgen induced KLK3 expression as expected, FOXA1 KD led to further increase in KLK3 level (Fig. 5b and Supplementary Fig. 11A). These increases could be blocked by anti-androgen bicalutamide, supporting the notion that AR transcriptional activity could be augmented upon FOXA1 loss. Analysis of ChIP-seq results indeed revealed a strong FOXA1-binding event within the intragenic region of the AR gene and this binding event was drastically reduced upon FOXA1 KD (Fig. 5c).

To determine whether this FOXA1-binding event leads to regulation of AR gene expression, QRT–PCR was carried out and showed that AR and KLK3 expression were indeed significantly increased following FOXA1 KD (Fig. 5d). Immunoblotting demonstrated remarkably upregulated KLK3 protein along with a slight increase of the AR protein. Similar results were also observed in other PCa cells including VCaP (Fig. 5e) and 22RV1 cells (Supplementary Fig. 11B). To preclude off-target shRNA effects, we used an siRNA that targets a different region of the FOXA1 gene and confirmed that FOXA1 KD indeed increased AR and KLK3 expression in PCa cells (Supplementary Fig. 11C,D). As western blot revealed only slight increase of AR protein upon FOXA1 KD, probably due to already very high level of AR in control cells, we took an alternative approach to overexpress FOXA1 in these cells to examine AR inhibition. Using FOXA1-expressing adenovirus, we demonstrated that FOXA1 overexpression remarkably suppressed both AR and KLK3 transcript and protein in PCa cells (Fig. 5f,g). Taken together, in addition to reducing specific AR binding by absorbing AR to excessive amount of open chromatin, FOXA1 also directly inhibited AR expression. This poses a pathologically important question as to whether FOXA1 loss in advanced PCa, as recently reported by us and others[15,20], might contribute to aberrant AR activation.

**FOXA1 loss results in androgen-independent AR binding events**. We first examined whether FOXA1 continues to inhibit AR in the absence of androgen. QRT–PCR and western blot analysis demonstrated that FOXA1 KD in hormone-deprived LNCaP cells increased AR expression and led to significant upregulation of KLK3 transcript as well as protein (Fig. 6a). In particular, although western blot was only able to detect slight increase in total AR protein, nuclear AR and chromatin-bound AR were both dramatically increased following the silencing of

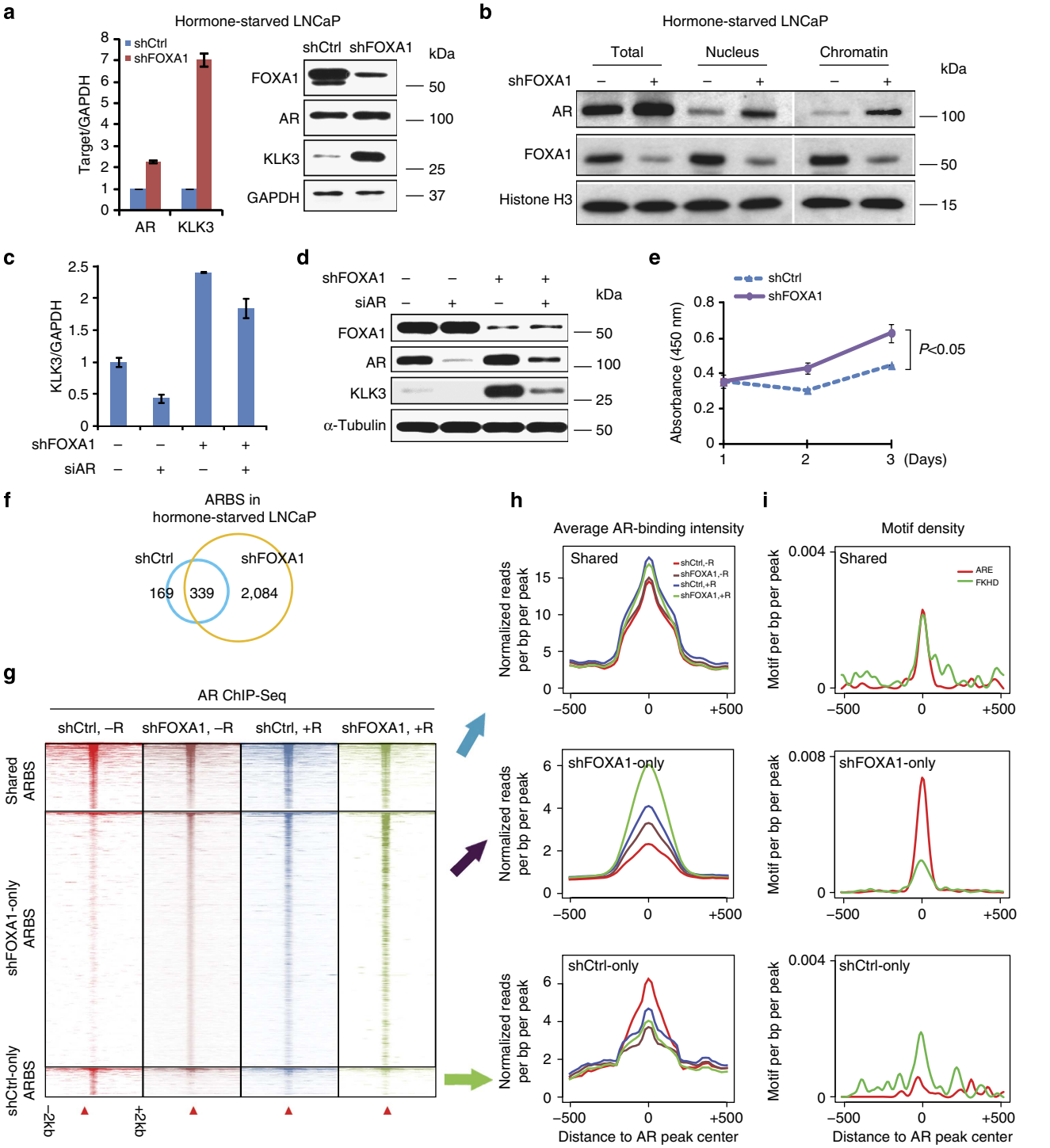



**Figure 6 | FOXA1 downregulation results in AR–chromatin binding in the absence of androgen.** (**a**) QRT–PCR and immunoblot analyses of AR, FOXA1 and KLK3 were performed in hormone-starved control and shFOXA1 LNCaP cells. GAPDH was used as a loading control. Error bars indicate $n=3$, mean ± s.e.m., $P<0.05$ by two-tailed Student's *t*-test. (**b**) Total lysate, nuclear proteins and chromatin fraction were collected or isolated from hormone-starved control and FOXA1-KD LNCaP cells and western blot was performed with indicated antibodies. (**c**,**d**) Control and shFOXA1 LNCaP cells were hormone deprived and subjected to RNA interference targeting control or AR (siAR). KLK3 gene expression was determined by QRT–PCR Error bars indicate $n=3$, mean ± s.e.m. $P<0.05$ by two-tailed Student's *t*-test. (**c**). Protein levels of KLK3, AR and FOXA1 were analysed by immunoblot (**d**). (**e**) Silencing of FOXA1 results in androgen-independent cell growth. LNCaP shCtrl and shFOXA1 cells were grown in hormone-deprived, phenol-free RPMI 1640 media supplemented with 5% charcoal-stripped fetal bovine serum. Total cell growth was measured by WST-1 assay as previously reported[20]. Error bars indicate $n=3$, mean ± s.e.m. Significant *P*-value was indicated. (**f**)AR ChIP-Seq was performed using hormone-deprived control and FOXA1-KD LNCaP cells. Venn diagram shows the overlap of ARBS between these two cell types. (**g**) Heatmap view of AR ChIP-seq-read density around ARBS (± 2 kb). ARBS detected in hormone-deprived shCtrl and shFOXA1 LNCaP cells were separated into shCtrl-only, shared and shFOXA1-only. '– R' indicated androgen-depleted medium, whereas '+R' indicated regular medium for LNCaP cell culture. (**h**) Average AR ChIP-seq intensity around ARBS (± 500 bp) shown in **f**. (**i**) ARE and FKHD motif density around ARBS (± 500 bp) shown in **f**.

FOXA1 (Fig. 6b). This suggests that loss of FOXA1 activated AR–chromatin binding in the absence of androgen. Indeed, ChIP–qPCR demonstrated remarkably increased AR binding on the KLK3 enhancer at a level comparable with androgen stimulation (Supplementary Fig. 12A). Concordantly, the expression levels of both KLK3 transcript and protein were drastically induced, which could be blocked by AR KD (Fig. 6c,d and Supplementary Fig. 12B). Most importantly, downregulation of FOXA1 significantly increased androgen-independent cell growth, reflecting the activation of the AR signalling pathway and its pathological relevance (Fig. 6e).To determine whether FOXA1 loss activates androgen-independent AR–chromatin binding globally, we mapped the genomic landscape of AR in the control and FOXA1-KD LNCaP cells that were grown in hormone-deprived medium. As expected, there were very few AR-binding events (508 ARBS) in the androgen-depleted control cells. Remarkably, following the silencing of FOXA1, a substantial amount of AR-binding events (over fourfold increase) was detected even in the absence of androgen (Fig. 6f).

We next examined how these androgen-independent ARBS relate to the normal androgen-stimulated AR-binding events. To do this, we categorized these ARBS into control-only, shFOXA1-only and shared, and examined these binding profiles in both androgen-deprived and androgen-stimulated conditions. Heatmap view of ChIP-seq-read intensity confirmed remarkable AR recruitment to the shFOXA1-only sites upon FOXA1 KD, despite a clear loss of FOXA1 binding (Fig. 6g and Supplementary Fig. 12C). Moreover, androgen stimulation was also able to recruit AR to these sites and showed a synergistic effect with FOXA1 KD (Fig. 6h). These results support that FOXA1 downregulation induces AR binding in the absence of androgen. For example, while a few AR binds the KLK3 enhancer in the control cells before androgen treatment, FOXA1 KD can lead to substantial AR recruitment, which is further enhanced by androgen stimulation (Supplementary Fig. 12D). Motif analysis revealed that the ARE motifs were strongly enriched in the ARBS gained upon FOXA1 KD (Fig. 6i and Supplementary Fig. 4A). Next, we investigated whether this increase in AR–chromatin binding by FOXA1 KD activates a corresponding AR transcriptional programme in the absence of androgen.

**FOXA1 loss activates a castration-resistant AR program**. To determine the AR-mediated transcriptional programme, we carried out microarray analysis of control and FOXA1-KD LNCaP cells grown in androgen-depleted medium. We separated FOXA1-regulated genes into FOXA1-induced and repressed and compared with those induced or repressed by androgen. Interestingly, GSEA analysis revealed that genes induced by FOXA1 in the absence of androgen significantly overlapped with those repressed by androgen, while FOXA1-repressed genes were enriched for androgen-induced genes (Fig. 7a,b). Therefore, in the absence of androgen, FOXA1 exhibited a transcriptional regulatory role that antagonizes AR. This is consistent with our ChIP-seq data indicating that FOXA1 KD induces AR–chromatin binding in the absence of androgen. Indeed, GSEA analysis showed that androgen-induced genes on the other hand were enriched for upregulation by FOXA1 KD, while androgen-repressed genes tended to become downregulated following the silencing of FOXA1 in hormone-deprived LNCaP cells (Fig. 7c,d). Furthermore, QRT–PCR analysis confirmed that known androgen-induced genes such as KLK3,TMPRSS2, FKBP5 and KLK2 were strongly induced by FOXA1 KD in hormone-deprived LNCaP cells, while previously reported androgen-repressed genes were inhibited (Figs 6c, 7e,f), in concordance with gained ARBS on main AR-binding loci of these genes upon FOXA1 depletion (Supplementary Fig. 13A–E). Therefore, FOXA1 loss induces AR transcriptional programme. In addition, gene ontology (GO) analysis of shFOXA1-induced gene set (Supplementary Data 1) revealed 'response to steroid hormone', 'regulation of cell proliferation' and 'locomotion/cell migration' as the enriched GO terms (Supplementary Fig. 14). This indicates that FOXA1 loss, like AR overexpression, provides important means to aberrant AR activation in the milieu of very low androgen.

## Discussion

Previous studies have established a model wherein FOXA1 protein acts as a pioneering factor that interacts with and recruits AR to lineage-specific binding sites[14]. Interestingly, two studies recently reported that FOXA1 KD resulted in extensive AR reprogramming rather than elimination, suggesting that FOXA1 may be pioneering for AR binding to some sites but inhibitory to other sites[15,16]. The mechanisms underlying this reprogramming, however, were not well understood. In the present study, we reconcile these conflicting results by providing a refined model wherein FOXA1 controls the genomic landscape of AR by altering chromatin accessibility, which may result in cooperative or inhibitory effects on AR programme depending on the relative FOXA1/AR abundance (Fig. 7g–i). FOXA1 is neither required nor sufficient to recruit AR binding. Rather, FOXA1 opens up chromatin to preferentially facilitate AR to access FOXA1-bound genomic regions. When AR is much more abundant than FOXA1 (for example, in LNCaP + shFOXA1 or DU145 + AR cells), AR primarily binds to its high-affinity target ARE sites across the genome (Fig. 7g). When FOXA1 reaches equilibrium with AR (for example, LNCaP cells), FOXA1 opens up FKHD-containing regions to permit AR interaction with either low-affinity hAREs or high-affinity full-AREs that are located within these regions, manifesting a collaborative role (Fig. 7h). This, however, is accomplished by removing AR from high-affinity ARE sites located outside of FXBS, as illustrated by the extensive loss of ARE-only ARBS. When FOXA1 becomes overly abundant (for example, hormone-deprived LNCaP cells or DU145 + AR + $FOXA1^{high}$ cells), FOXA1 opens up excessive amount of genomic regions, which act as a reservoir that absorbs AR through low-affinity interactions and thus drastically reduces the amount of AR available for any individual loci. In this condition, FOXA1 inhibits almost all specific AR–chromatin binding (Fig. 7i). This context-dependent regulation may be critical in offsetting androgen response in hormone-deprived condition and in maximizing rapid response upon androgen stimulation, which should be carefully examined in future studies. Our proposed model demonstrates multifaceted roles of FOXA1 in tightly controlling AR signalling and provides mechanisms that reconcile previously reported roles of FOXA1 as an AR-collaborating or inhibiting co-factor[14–16].

Importantly, we believe that our model precisely recapitulates how FOXA1 dictates prostatic-specific gene expression. Had FOXA1 not affected AR–chromatin binding, AR would have been relinquished to bind ARE motifs across the genome, which are not specific to prostate genes. On the other hand, had FOXA1 directly recruited AR to the genome AR would potentially get recruited to all FXBS, which would ultimately dilute AR and inhibits AR-mediated activation of prostate-specific genes. Our model, by contrast, showed that when FOXA1 is in equilibrium with AR, it opens up the chromatin to preferably facilitate AR co-occupancy at FKHD + ARE sites (high-affinity and highly accessible), but reduces AR binding to ARE-only sites (high affinity but not accessible). Therefore, our study not only supports previously reported roles of FOXA1 in specifying lineage-specific AR programme[14], but also delineates the

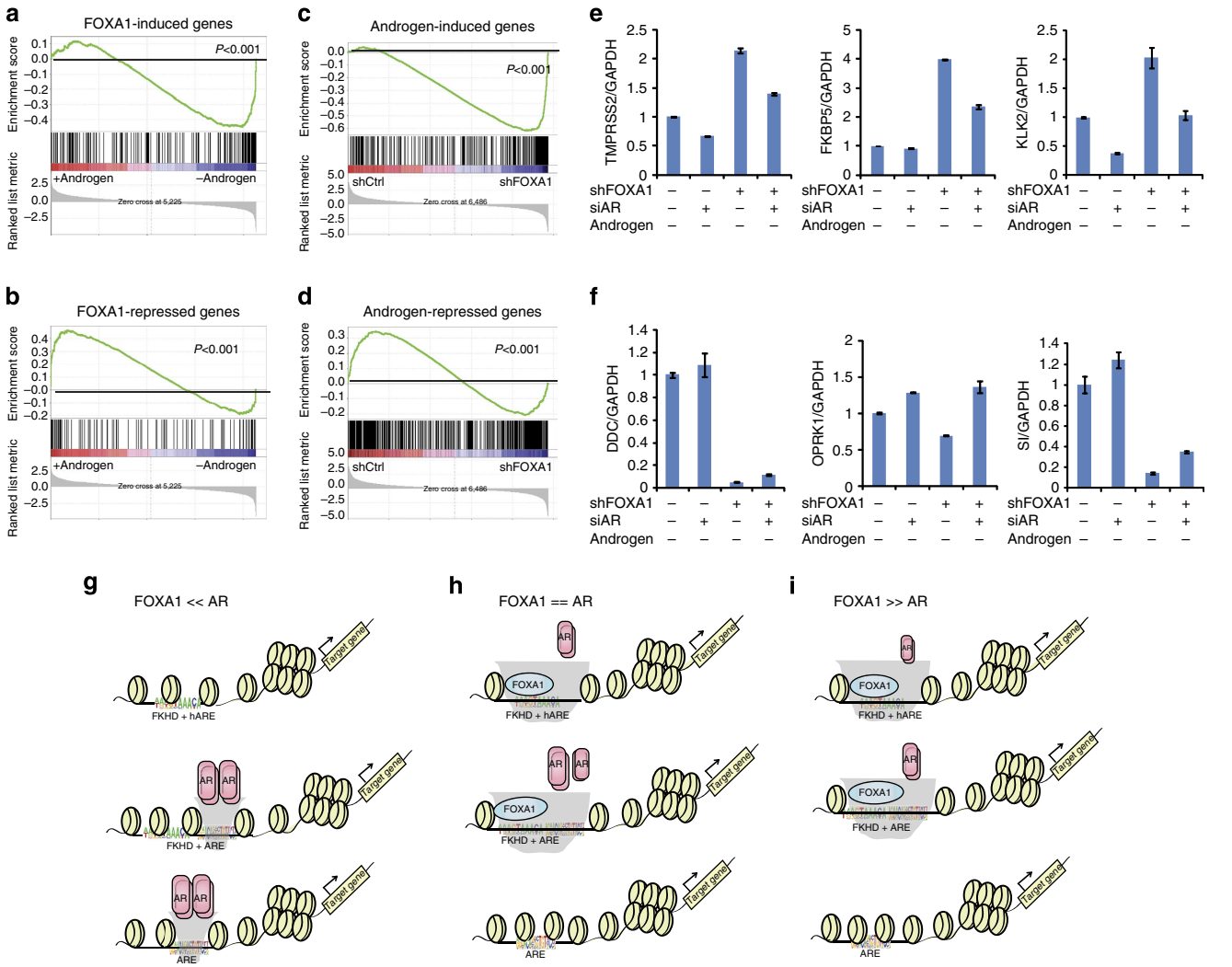



**Figure 7 | FOXA1 loss activates a castration-resistant AR transcriptional programme.** (**a**,**b**) GSEA was carried out using FOXA1-induced or FOXA1-repressed gene set comparing control and androgen-treated LNCaP cells. Genes induced by FOXA1 in hormone-starved LNCaP cells are enriched for repression by androgen (**a**). Genes repressed by FOXA1 in hormone-starved LNCaP cells are enriched for induction by androgen (**b**). FOXA1-induced or repressed gene set (with 2-fd cutoff) was obtained from microarray profiling control and shFOXA1 LNCaP cells grown in hormone-starved condition. Microarray data have been previously released at GEO under accession GSE37314 (ref. 20). (**c**,**d**) GSEA reveals that androgen-induced genes correlate with higher expression in shFOXA1 cells relative to control cells (**c**), while androgen-repressed genes are significantly enriched for downregulation in shFOXA1 cells (**d**). Androgen-induced and -repressed genes were identified from microarrays of LNCaP shCtrl –R cells and LNCaP shCtrl + R cells using a cutoff of ≥twofold change. Expression data set are from microarray profiling of hormone-starved LNCaP shCtrl and shFOXA1 cells as above mentioned. (**e**,**f**) QRT-PCR analysis of expression of three typical androgen-induced genes (**e**) and three androgen-repressed genes (**f**) in hormone-starved LNCaP cells treated with shFOXA1, siAR or both. All error bars indicate $n = 3$, mean ± s.e.m. (**g**–**i**) Model of FOXA1 regulation of AR–chromatin binding in a dose- and context-dependent manner. (**g**) When FOXA1 is relative low or absent, AR is freely available to bind genomic regions containing ARE. (**h**) When FOXA1 and AR levels are in equilibrium, FOXA1 opens up FKDH-containing chromatin regions to facilitate AR binding to these sites, thereby preventing AR binding to ARE-only sites. (**i**) When FOXA1 is much more abundant than AR, it opens up excessive amount of chromatin regions, which end up diluting AR across the genome, thereby inhibiting specific AR-binding events. Cup-shape shaded areas indicate chromatin accessibility. hARE indicates low-affinity ARBS with hARE motif, compared with high-affinity ARBS with canonical ARE motif.

mechanistic details of this regulation. The ability of FOXA1 to inhibit high-affinity ARE-mediated AR–chromatin binding is not only essential in defining a prostatic AR programme but also critical in preventing non-prostatic AR-binding events. Recent studies have shown that such programme could be highly oncogenic[15] and alteration of this regulatory pathway may turn on novel AR target genes in CRPC[21]. Future studies should examine how the balanced regulation between FOXA1 and AR is disrupted during PCa progression.

In this study, we have also shown that FOXA1 directly inhibits AR expression and thus the transcription of its target genes. Interestingly, consistent with our results, a recent study has shown that FOXA1 overexpression largely inhibits AR target genes in PCa cells[22] and decreases cell motility and tumour metastasis[15,20]. Most importantly, our study revealed that FOXA1 loss activates AR–chromatin binding and AR transcriptional regulation in the absence of androgen. This activation may be due to the fact that FOXA1 KD enabled the cells to reach a new balance between reduced FOXA1 and the low amount of nuclear AR available in hormone-deprived cells, thereby permitting high-affinity AR binding. In addition, cell growth assay showed increased PCa cell growth following the silencing of FOXA1 in

androgen-depleted medium. These findings suggest that loss of FOXA1 may lead to androgen-independent AR signalling and thus CRPC progression. Indeed, we have recently reported that FOXA1 is downregulated in CRPC[20]. Due to the multifaceted roles of FOXA1 in regulating AR signalling that is heavily dependent on FOXA1-AR equilibrium, the pathological relevance of FOXA1 may be elusive. In fact, there are apparent conflicting reports that linked FOXA1 expression levels with either good or bad clinical outcomes. Based on our model, it is possible that FOXA1 may be associated with poor prognosis if the analysis is solely based on patients with primary PCa[16] or metastatic PCa[23] who have not received prior systematic therapy. However, FOXA1 may be a good outcome predictor in patient cohorts with prior androgen-deprivation therapy[15]. It is thus very important to re-evaluate FOXA1 association with PCa outcomes in future studies by carefully stratifying patients based on their disease stage and treatment histories. Taken together, our study delineates the mechanism by which FOXA1 defines the genomic action of AR and provides a novel pathway to aberrant AR activation in the milieu of very low androgen.

## Methods

**Cell Culture and antibodies.** LNCaP, DU145, 22RV1 and VCaP cells were obtained from the American Type Culture Collection (ATCC) and grown in RPMI or DMEM (VCaP) supplemented with 10% fetal bovine serum and standard antibiotics. LAPC4 cells were provided by Dr C Shad Thaxton (Northwestern University) and grown in IMEM supplemented with 10% fetal bovine serum and 1 nm R1881. For androgen treatment, cells were hormone starved for 2–3 days and treated with 1 nM R1881 for 24 h. Bicalutamide (Casodex) was added at a concentration of 10 μM as needed. Antibodies used in this study include anti-AR (06-680) from Millipore, anti-FOXA1 (ab23738), Histone H3 (ab1791) and anti-GAPDH (ab9385) from Abcam, anti-AR (sc-7305) from Santa Cruz and anti-KLK3 (K2889-100UG) from Sigma (Supplementary Table 3).

**siRNA, shRNA and Plasmids.** FOXA1 siRNA (sense 5′-GAGAGAAAAAAUCAA CAGC-3′; antisense 5′-GCUGUUGAUUUUUUCUCUC-3′)[14,20], AR on-target plus smart pool (L-003400-00-0020), and control siRNA Luciferase GL2 Duplex (D-001100-01-20) were synthesized by Dharmacon. The control and pGIPZ lentiviral shRNAmir targeting FOXA1 (Clone ID# V2LHS_16780) were purchased from Open Biosystems. C-terminal 3 × HA-tagged full-length FOXA1 was cloned into a Tet-On inducible lentiviral vector that was kindly provided by Professor Junjie Chen (MD Anderson Cancer Center). The full-length (wt) and truncated AR (Q640X) were amplified from pcDNA3.1-AR$^{fl}$ plasmid[24] and cloned into the pCR8/GW/TOPO entry vector (Invitrogen). The pCR8-AR C562S and G568W mutants were generated from pCR8-AR wt using QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA). Lentiviral constructs were generated by LR recombination between pCR8-AR constructs and pLenti CMV/TO Puro DEST (Addgene plasmid 17293). Details regarding cloning primers and plasmid construction were provided in Supplementary Tables 1 and 2.

**Western blot analysis.** Cells were lysed in NP40 Cell Lysis Buffer (Invitrogen). Protein concentration was measured using the Bio-Rad protein assay kit. Twenty micrograms of the total protein were mixed with SDS sample buffer, boiled for 10 min at 95 °C, separated on a 10% SDS-polyacrylamide gel and transferred to an Amersham Hybond PVDF membrane. The membranes were blocked with 5% w/v BSA or milk in TBST overnight at 4 °C, incubated in primary antibody (diluted in blocking solution) for either 1 h at room temperature or overnight at 4 °C (for antibodies, see Supplementary Table 3), washed 3 × with TBST and incubated for 1 h in a secondary antibody diluted in blocking solution. Membranes were again washed 3 × with TBST and ECL (GE Healthcare) was added to the membrane and the signal was detected on film (GE Healthcare). Representative uncropped raw scans of blots generated with the most relevant antibodies used along the different figures are present in Supplementary Figs 15–17.

**Co-immunoprecipitation.** Cells with coexpression of FOXA1 (tagged with Flag) and AR (wt or mutant) were lysed in IP lysis buffer (50 mM Tris-Cl pH7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, Roche protease inhibitor cocktail). An aliquot of the cell lysate was kept as input for western blot analysis. Cell lysate was first precleared with protein A/G-agarose beads at 4 °C for 2 h. Then the precleared lysate was incubated with mouse anti-Flag (F1804, Sigma) or mouse anti-AR antibody (sc-7305, Santa Cruz), mouse IgG (as a negative control) overnight on a rocker platform. The next day, protein A/G-agarose beads were added to the mixture and incubated at 4 °C for 2 h. Then the beads were pelleted and washed for four times with IP lysis buffer before boiling in SDS sample buffer for 10 min. Western blotting analysis were performed using rabbit anti-FOXA1(ab23738, Abcam) or rabbit anti-AR(06-680, Millipore).

**Chromatin fractionation.** Chromatin was isolated as described with the following modifications[25]. Cells were resuspended in buffer A (10 mM HEPES, [pH 7.9], 10 mM KCl, 1.5 mM MgCl2, 0.34 M sucrose, 10% glycerol, 1 mM DTT, 1 × Roche protease inhibitor cocktail). Triton X-100 (0.1%) was added, and the cells were incubated for 5 min on ice. Nuclei were collected in pellet 1 by low-speed centrifugation (5 min, 1,300 *g*, 4 °C), washed once with buffer A, and then resuspended in buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT, protease cocktail) for 30 minutes on ice. Insoluble chromatin was isolated by centrifugation (5 min, 1,700 *g*, 4 °C), washed once with buffer B, and centrifuged again under the same conditions. The final chromatin pellet was resuspended in SDS sample buffer (50 mM Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 2% β-mercaptoethanol and 0.02% bromophenol blue).

**Chromatin immunoprecipitation (ChIP).** AR and FOXA1 ChIP were carried out as described previously[9,26,27]. Briefly, cultured cells were cross-linked with 1% formaldehyde for 10 min and the cross-linking was inactivated by 0.125 M glycine for 5 min at room temperature. Cells were then rinsed with cold 1 × PBS twice. The following steps were performed at 4 °C. Cell pellets were resuspended and incubated in cell lysis buffer + 10 ul ml$^{-1}$ PMSF and protease inhibitor (Roche) for 10 min. Nuclei pellets were spinned down at 5,000 *g* for 5 min, resuspended in nuclear lysis buffer and then incubated for another 10 min. Chromatin was sonicated to an average length of 500 bp and then centrifuged at 14,000 *g* for 10 min to remove the debris. Supernatants containing chromatin fragments were incubated with agarose/protein A or G beads (Upstate) for 15 min and centrifuged at 5,000 *g* for 5 min to reduce nonspecific binding. To immunoprecipitate protein/chromatin complexes, the supernatants were incubated with 3–5 ug of antibody (anti-FOXA1 from Abcam, cat# ab23738; anti-AR from Millipore, cat#06-680) overnight, then added 50 ul of agarose/protein A or G beads and incubated for 1.5 h. Beads were washed twice with 1X dialysis buffer and four times with IP wash buffer. The antibody /protein/DNA complexes were eluted with 150 ul IP elution buffer twice. To reverse the cross-links, the complexes were incubated in elution buffer + 10 ug RNase A and 0.3 M NaCl at 67 °C for 4 h. DNA/proteins were precipitated with ethanol, air-dried and dissolved in 100 ul of TE. Proteins were then digested by proteinase K at 45 °C for 1 h and DNA was purified with QIAGEN PCR column and eluted with 30 ul EB.

**Formaldehyde-assisted isolation of regulatory elements.** FAIRE was performed according to the protocol published by Giresi *et al.*[28] Briefly, cells were cross-linked identically as for ChIP. After 10-min cross-linking with 1% formaldehyde and stopping with 0.125 M glycine, the washed cell pellets (10 e7 cells) were resuspended and incubated for 10 min sequentially in 1 ml of buffer L1, 1 ml of buffer L2 and 400 μl of buffer L3. The lysates were sonicated in order to shear chromatin into DNA fragments of average 300–500 bp, then centrifuged at 14,000 *g* for 10 min to remove the cellular debris. Input samples were reverse cross-linked overnight at 65 °C. The FAIRE samples and reverse cross-linked input samples were subjected to three sequential phenol/chloroform/isoamyl alcohol (25/24/1) extractions. DNA were precipitated with ethanol, air-dried and dissolved in 100 ul of TE and treated with 1 μl of RNase A (10 mg ml$^{-1}$) for 1 h at 37 °C. Proteins were then digested by proteinase K at 45 °C for 1 h and the DNA was purified with QIAquick PCR Purification Kit (Qiagen) and eluted with 30 ul EB.

**Quantitative Polymerase Chain Reaction.** All primers (Supplementary Table 1) were designed using Primer 3 (http://frodo.wi.mit.edu/primer3/), and synthesized by Integrated DNA Technologies. SYBR Green based quantitative real-time PCR was performed using GoTaq qPCR MasterMix (Promega) using a StepOnePlus Real-Time PCR System (Applied Biosystems). For quantitative reverse transcription-PCR (QRT–PCR) data analysis, the fold change in the target gene relative to the GAPDH (glyceraldehyde 3-phosphate dehydrogenase) endogenous control gene is determined by: fold change $= 2^{-\Delta(\Delta Ct)}$ where $\Delta Ct = Ct_{target} - Ct_{GAPDH}$ and $\Delta(\Delta Ct) = \Delta Ct_{treatment} - \Delta Ct_{control}$. For ChIP–qPCR and FAIRE–qPCR, enrichment analysis were performed by Comparative Ct method and normalization to input, that is, enrichment over input $= 2^{(-\Delta Ct)}$, where $\Delta Ct = Ct_{sample} - Ct_{input}$.

**ChIP-seq and FAIRE-seq library preparation and sequencing.** Libraries were prepared according to standard protocols using BioScientific's DNA Sample Kit (Cat# 514101)[9,27]. Libraries were sequenced using Illumina Hi-Seq platforms. Sequence reads were aligned to the Human Reference Genome (assembly hg19) using Burrows-Wheeler Alignment (BWA) Tool version 0.6.1 (ref. 29).

**ChIP-seq and FAIRE-seq peak discovery.** ChIP-seq and FAIRE-seq peak identification, overlap, subtraction, union and feature annotation of enriched regions were performed using HOMER (Hypergeometric Optimization of Motif EnRichment) suite (http://homer.salk.edu/homer/)[30]. For ChIP-seq, HOMER peak finding algorithm searches for peaks of fixed size to achieve maximum sensitivity.

For FAIRE-seq analysis, we used HOMER to find variable-width enriched regions. Homer automatically optimized parameters needed for peak finding the size of the peak and the length of fragments. We set the minimum distance between peaks using 2.5 × the peak size. The cutoff for statistically significant peaks was selected at a false discovery rate of 0.001 determined by peak finding using randomized tag positions in a genome with an effective size of 2 billion basepairs. We required the tag density at peaks to be fourfold greater than in the surrounding 10 kb region to exclude putative peaks identified in regions of genomic duplication. Only one tag from each unique position was considered to filter clonal signal from the sequencing. Additional peak calls were determined using MACS peak caller version 1.4.1 (ref. 31) with default parameters, in which enriched regions of the genome were identified by comparing the ChIP samples to input samples. The number of sequencing reads and AR or FOXA1-binding events for each sample were shown in Supplementary Table 4.

**Motif discovery.** The HOMER motif discovery was used to perform *de novo* motif discovery as well as check the enrichment of known motifs in a set of given genomic region (200 bp surrounding ChIP-seq peak center)[30]. Motif density histograms were created using HOMER for target regions. Control regions were generated by selecting DNA sequences of equal length at 10 kb downstream of the target regions. The motif density at target regions was normalized to that at the control regions. Additional motif analysis was determined using MEME-ChIP algorithm[32].

**Gene expression array and data analysis.** Total RNAs were isolated using TRIzol reagent (Invitrogen). The integrity of the RNA was monitored using Bioanalyzer 2100. Microarray profiling was performed using HumanHT-12 v 4.0 Expression BeadChip (Illumina). Bead-level data were preprocessed using GenomeStudio (Illumina), and the expression values were quantile-normalized using the bead array package in Bioconductor. Differentially expressed genes were identified using a twofold cutoff. GO terms enrichment was analysed using PANTHER (http://www.pantherdb.org/)[33] or GOrilla (http://cbl-gorilla.cs.technion.ac.il/)[34]. GSEA[35] was performed as described previously[36]. Gene expression fold changes in experimental cells relative to control were calculated and genes having at least twofold changes in shFOXA1-treated cells were defined as FOXA1-regulated gene sets. Differential expressed genes (at least twofold changes) in LNCaP shCtrl cells androgen-treated compared with hormone-deprived were defined as androgen-regulated gene sets. Raw and normalized data from microarray experiments have been deposited at NCBI Gene Expression Omnibus database (under accession GSE37314 and GSE55007).

**Cell proliferation assay.** Cell proliferation assay was carried out using the WST-1 kit according to the manufacturer's instruction (Clontech). Briefly, cells were seeded in a 24-well plate at a density of 10,000 cells in 500 ul of complete culture medium and cultured in a $CO_2$ incubator at 37 °C for 24 h. Then cells were hormone starved for 2 to 3 days prior to WST-1 assay. Cultures were incubated for 2 h after the addition of the premixed WST-1 cell proliferation reagent and the absorbance at 450 nm was measured using a multiwell plate reader. The absorbance level of samples was normalized to that of the background control well (containing culture medium plus WST-1 reagent, without cells).

## References


1. Berger, R. *et al.* Androgen-induced differentiation and tumorigenicity of human prostate epithelial cells. *Cancer Res.* **64,** 8867–8875 (2004).
2. Heinlein, C. A. & Chang, C. Androgen receptor in prostate cancer. *Endocr. Rev.* **25,** 276–308 (2004).
3. Heemers, H. V. & Tindall, D. J. Androgen receptor (AR) coregulators: a diversity of functions converging on and regulating the AR transcriptional complex. *Endocr. Rev.* **28,** 778–808 (2007).
4. Balk, S. P. & Knudsen, K. E. AR, the cell cycle, and prostate cancer. *Nucl. Recept. Signal.* **6,** e001 (2008).
5. Lamont, K. R. & Tindall, D. J. Minireview: alternative activation pathways for the androgen receptor in prostate cancer. *Mol. Endocrinol.* **25,** 897–907 (2011).
6. Yuan, X. & Balk, S. P. Mechanisms mediating androgen receptor reactivation after castration. *Urol. Oncol.* **27,** 36–41 (2009).
7. Wang, Q. *et al.* A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. *Mol. Cell* **27,** 380–392 (2007).
8. Lin, B. *et al.* Integrated expression profiling and ChIP-seq analyses of the growth inhibition response program of the androgen receptor. *PLoS ONE* **4,** e6589 (2009).
9. Yu, J. *et al.* An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell* **17,** 443–454 (2010).
10. Cirillo, L. A. *et al.* Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9,** 279–289 (2002).
11. Gao, N. *et al.* Forkhead box A1 regulates prostate ductal morphogenesis and promotes epithelial cell maturation. *Development* **132,** 3431–3443 (2005).
12. Gao, N. *et al.* The role of hepatocyte nuclear factor-3 alpha (Forkhead Box A1) and androgen receptor in transcriptional regulation of prostatic genes. *Mol. Endocrinol.* **17,** 1484–1507 (2003).
13. Yu, X. *et al.* Foxa1 and Foxa2 interact with the androgen receptor to regulate prostate and epididymal genes differentially. *Ann. NY Acad. Sci.* **1061,** 77–93 (2005).
14. Lupien, M. *et al.* FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132,** 958–970 (2008).
15. Wang, D. *et al.* Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474,** 390–394 (2011).
16. Sahu, B. *et al.* Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. *EMBO J.* **30,** 3962–3976 (2011).
17. Shim, E. Y., Woodcock, C. & Zaret, K. S. Nucleosome positioning by the winged helix transcription factor HNF3. *Genes Dev.* **12,** 5–10 (1998).
18. Sekiya, T., Muthurajan, U. M., Luger, K., Tulin, A. V. & Zaret, K. S. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. *Genes Dev.* **23,** 804–809 (2009).
19. He, H. H. *et al.* Nucleosome dynamics define transcriptional enhancers. *Nat. Genet.* **42,** 343–347 (2010).
20. Jin, H. J., Zhao, J. C., Ogden, I., Bergan, R. C. & Yu, J. Androgen receptor-independent function of FoxA1 in prostate cancer metastasis. *Cancer Res.* **73,** 3725–3736 (2013).
21. Wang, Q. *et al.* Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell* **138,** 245–256 (2009).
22. Grasso, C. S. *et al.* The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487,** 239–243 (2012).
23. Jain, R. K., Mehta, R. J., Nakshatri, H., Idrees, M. T. & Badve, S. S. High-level expression of forkhead-box protein A1 in metastatic prostate cancer. *Histopathology* **58,** 766–772 (2011).
24. Sun, S. *et al.* Castration resistance in human prostate cancer is conferred by a frequently occurring androgen receptor splice variant. *J. Clin. Invest.* **120,** 2715–2730 (2010).
25. Mendez, J. & Stillman, B. Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. *Mol. Cell Biol.* **20,** 8602–8612 (2000).
26. Wu, L. *et al.* CCN3/NOV gene expression in human prostate cancer is directly suppressed by the androgen receptor. *Oncogene* **33,** 504–513 (2014).
27. Zhao, J. C. *et al.* Cooperation between polycomb and androgen receptor during oncogenic transformation. *Genome Res.* **22,** 322–331 (2012).
28. Giresi, P. G. & Lieb, J. D. Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (formaldehyde assisted isolation of regulatory elements). *Methods* **48,** 233–239 (2009).
29. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
30. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38,** 576–589 (2010).
31. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
32. Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA data sets. *Bioinformatics* **27,** 1696–1697 (2011).
33. Thomas, P. D. *et al.* PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* **13,** 2129–2141 (2003).
34. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10,** 48 (2009).
35. Subramanian, A., Kuehn, H., Gould, J., Tamayo, P. & Mesirov, J. P. GSEA-P: a desktop application for gene set enrichment analysis. *Bioinformatics* **23,** 3251–3253 (2007).
36. Yu, J. *et al.* Integrative genomics analysis reveals silencing of beta-adrenergic signaling by polycomb in prostate cancer. *Cancer Cell* **12,** 419–431 (2007).


## Acknowledgements

We thank Chris Runkle for technical supports and Angela Yang for helpful discussion. This work was supported in part by the NIH P50CA090386 pilot project (to J.Y.), U54CA143869 pilot project (to J.Y.), R01CA172384 (to J.Y.), the U.S. Department of Defence W81XWH-13-1-0319 (to J.Y.), and the Research Scholar Award RSG-12-085-01 (to J.Y.) from the American Cancer Society.

## Author contributions

J.Y. and H.-J.J. conceived the study and designed the experiments. H.-J.J. performed a majority of the experiments with assistance from L.W. and J.K.; L.W. generated ChIP-seq libraries; J.C.Z and H.-J.J. analyzed the ChIP-seq and microarray data with guidance from J.Y.; J.K. facilitated with graphic illustration. J.Y. and H.-J.J. wrote the paper. All authors discussed the results and commented on the manuscript.

## Additional information

**Accession codes**: New high-throughput sequence and microarray data generated in this study have been deposited in GEO database under accession code GSE55007.

**Supplementary Information** accompanies this paper on http://www.nature.com/naturecommunications

**Competing financial interests**: The authors declare no competing financial interests.

**Reprints and permission** information is available online at http://npg.nature.com/reprintsandpermissions/

**How to cite this article**: Jin, H.-J. *et al.* Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional programme. *Nat. Commun.* 5:3972 doi: 10.1038/ncomms4972 (2014).

Oncogene (2015), 1–10

www.nature.com/onc

ORIGINAL ARTICLE

# LncRNA *HOTAIR* enhances ER signaling and confers tamoxifen resistance in breast cancer

X Xue[1,2,4], YA Yang[2,4], A Zhang[2], K-W Fong[2], J Kim[2], B Song[2], S Li[2], JC Zhao[2] and J Yu[2,3]

Tamoxifen, an estrogen receptor (ER) antagonist, is the mainstay treatment of breast cancer and the development of resistance represents a major obstacle for a cure. Although long non-coding RNAs such as HOTAIR have been implicated in breast tumorigenesis, their roles in chemotherapy resistance remain largely unknown. In this study, we report that *HOTAIR* (HOX antisense intergenic RNA) is upregulated in tamoxifen-resistant breast cancer tissues compared to their primary counterparts. Mechanistically, HOTAIR is a direct target of ER-mediated transcriptional repression and is thus restored upon the blockade of ER signaling, either by hormone deprivation or by tamoxifen treatment. Interestingly, this elevated *HOTAIR* increases ER protein level and thus enhances ER occupancy on the chromatin and potentiates its downstream gene regulation. *HOTAIR* overexpression is sufficient to activate the ER transcriptional program even under hormone-deprived conditions. Functionally, we found that *HOTAIR* overexpression increases breast cancer cell proliferation, whereas its depletion significantly impairs cell survival and abolishes tamoxifen-resistant cell growth. In conclusion, the long non-coding RNA *HOTAIR* is directly repressed by ER and its upregulation promotes ligand-independent ER activities and contributes to tamoxifen resistance.

*Oncogene* advance online publication, 14 September 2015; doi:10.1038/onc.2015.340

## INTRODUCTION

Long non-coding RNAs (lncRNAs) are a major class of newly identified non-coding transcripts that are usually composed of more than 200 nucleotides. Accumulating evidence suggests that lncRNAs play critical roles in regulating a wide range of cellular processes by affecting various aspects of protein, DNA, and RNA expression and interactions.[1–5] Large-scale RNA sequencing (RNA-seq) studies have revealed that lncRNAs are abundantly transcribed from the genome; a recent study comprehensively examined over 7000 RNA-seq libraries and uncovered nearly 60,000 lncRNAs from the human transcriptome.[6] Out of numerous cancer-associated lncRNAs, *HOTAIR* (HOX antisense intergenic RNA) was among the most upregulated in breast cancer. Localized in chromosome 12, *HOTAIR* is 2.2 kb in length and transcribed from the antisense strand of the *HOXC* locus. It has been shown to interact with polycomb repressive complex 2 to reprogram the chromatin state and induce cancer metastasis.[7,8] *In vivo* experiments showed that *HOTAIR* is sufficient and required to promote invasion of breast carcinoma cells.[7] Concordantly, *HOTAIR* and EZH2 expression levels were highly correlated in breast cancer tissues and high *HOTAIR* level is associated with worse prognosis.[9,10] In addition, these studies reported that strong HOTAIR expression correlated with estrogen receptor (ER) and PR positivity, and that *HOTAIR* expression is a strong predictor of poor clinical outcome especially in ER-positive breast cancer.[9,10]

These results provide first lines of evidence that the lncRNA *HOTAIR* may play important roles in regulating breast cancer progression. Tamoxifen, an antagonist of the ER, is the most commonly used treatment for ER-positive breast cancer. Despite great success in improving the overall survival of breast cancer patients, development of tamoxifen resistance (TamR) is persistently seen in the clinic and is a major cause of breast cancer recurrence and mortality.[11] Understanding the biological mechanisms underlying this acquired resistance to tamoxifen is thus of substantial clinical significance.[12]

ER is a hormonal transcription factor that is liganded and activated by estrogen. ER regulates target genes that control endocrine response and cell cycle progression.[5,13,14] Tamoxifen competes with estrogen for binding to the ER protein, thereby inhibiting the conventional ER transcriptional program.[5,14,15] Using chromatin immunoprecipitation sequencing (ChIP-seq), a recent study has mapped genome-wide ER binding profiles in primary breast cancers and found that ER is still recruited to the chromatin in TamR breast cancer, but to new regulatory regions associated with poor clinical outcome.[16] This aberrant ER transcriptional activity is proposed to be regulated by various oncogenic mechanisms and have critical functions in mediating tamoxifen resistance and tumor progression. Here we report that *HOTAIR* is overexpressed in TamR breast cancer. It directly interacts with the ER protein to enhance ER transcriptional activity and thus ligand-independent breast cancer growth. Our study will not only inform about the mechanistic underpinnings of breast cancer progression but also provide evidence supporting therapeutic potentials of lncRNA targeting in breast cancer treatment.

## RESULTS

### *HOTAIR* is upregulated in tamoxifen-resistant, ER-positive breast cancer

To determine lncRNAs that may contribute to breast cancer tamoxifen resistance, we re-analyzed publicly available data set

[1]Division of Thoracic Surgery, Cancer Center of Guangzhou Medical University, Guangzhou, Guangdong, China; [2]Division of Hematology/Oncology, Department of Medicine, Robert H. Lurie Comprehensive Cancer Center, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA and [3]Robert H. Lurie Comprehensive Cancer Center, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA. Correspondence: Dr J Yu, Division of Hematology/Oncology, Department of Medicine, Robert H. Lurie Comprehensive Cancer Center, Northwestern University, Feinberg School of Medicine, 303 E. Superior St. Lurie 5-117, Chicago, IL 60611, USA.
E-mail: jindan-yu@northwestern.edu
[4]These authors contributed equally to this work.
Received 8 May 2015; revised 23 July 2015; accepted 28 July 2015

profiling gene expression in wild-type MCF7 cells as well as its TamR derivatives treated with ethanol or 17β-estradiol (E2) for 4 h (GSE5840).[17] Our analysis revealed 37 lncRNA genes that were repressed by estrogen and became upregulated in TamR cells (Figure 1a). Among the top deregulated lncRNAs are *HOTAIR* and TP53TG1. Although *HOTAIR* has been shown to be upregulated in metastatic breast cancer,[7,10] its role in TamR has not been investigated. To examine this, we performed *in situ* hybridization to probe the abundance of *HOTAIR* lncRNA in breast cancer tissues, comparing between matched primary and TamR breast carcinoma samples. Our results showed that *HOTAIR* localized primarily in the nuclei but was also present in the cytoplasm (Figure 1b). Most primary breast cancer tissues had weak *HOTAIR* staining, whereas TamR breast cancer generally exhibited moderate to strong *HOTAIR* staining. Overall, *HOTAIR* expression level was significantly higher in TamR breast cancer than primary, hormone-naive tumors (Figure 1c). Being consistent with this, quantitative reverse transcriptase – PCR analysis showed that tamoxifen treatment for 7 days significantly increased *HOTAIR* lncRNA levels in both MCF7 and T47D cells, while dramatically decreasing the expression of *GREB1*, a known ER-induced gene (Figures 1d and e). As tamoxifen is known to compete with estrogen to inhibit estrogen-induced ER activities, next we examined whether *HOTAIR* is a target of ER-mediated transcriptional regulation.

## The lncRNA *HOTAIR* is directly repressed by estrogen receptor

To examine whether estrogen regulates *HOTAIR* expression, we carried out quantitative reverse transcriptase – PCR analysis of MCF7 cells treated with increasing doses of E2. *HOTAIR* expression was greatly inhibited for up to sevenfold, while *GREB1* was increased as expected (Figure 2a and Supplementary Figure S1A). Estrogen inhibited *HOTAIR* expression in a dose- and time-dependent manner (Figures 2a and b). *HOTAIR* level was decreased about twofold after 4 h of E2 treatment and nearly 10-fold after 24 h of E2 treatment, whereas *GREB1* was gradually induced by around 20-fold at 4 h and reached a plateau of >30-fold after 8 h (Supplementary Figure S1B). A similar trend of inhibition of *HOTAIR* expression by estrogen was observed in a different ER+ breast cancer cell line T47D, despite T47D being much less responsive to estrogen as indicated by much less *GREB1* induction (Figures 2c and d). Furthermore, *HOTAIR* level is considerably restored in breast cancer cells following hormone deprivation, wherein *GREB1* expression was lost (Figure 2e). Next, to determine whether estrogen inhibits *HOTAIR* expression through direct ER binding to *HOTAIR* regulatory elements, we re-analyzed a previously published study involving an ER ChIP-seq data set that was performed in MCF7 cells (GSE23893).[18] We observed a very strong ER binding site at a genomic region about 14.5 kb upstream to the transcription start site of the *HOTAIR* gene (Figure 2f). In addition, this region is strongly occupied by H3K4me1 and H3K27ac (GSE40129), supporting its being an active enhancer (Supplementary Figure S2A). ER ChIP followed by quantitative reverse transcriptase – PCR analysis confirmed that estrogen stimulation significantly increased ER binding to this region as well as to positive control gene *GREB1*, but not to the negative control gene *KIAA0066* (Figure 2g). Further, chromosome conformation capture (3C) experiment demonstrated estrogen-induced DNA looping between the transcription start site of the *HOTAIR* gene (anchor primer) and

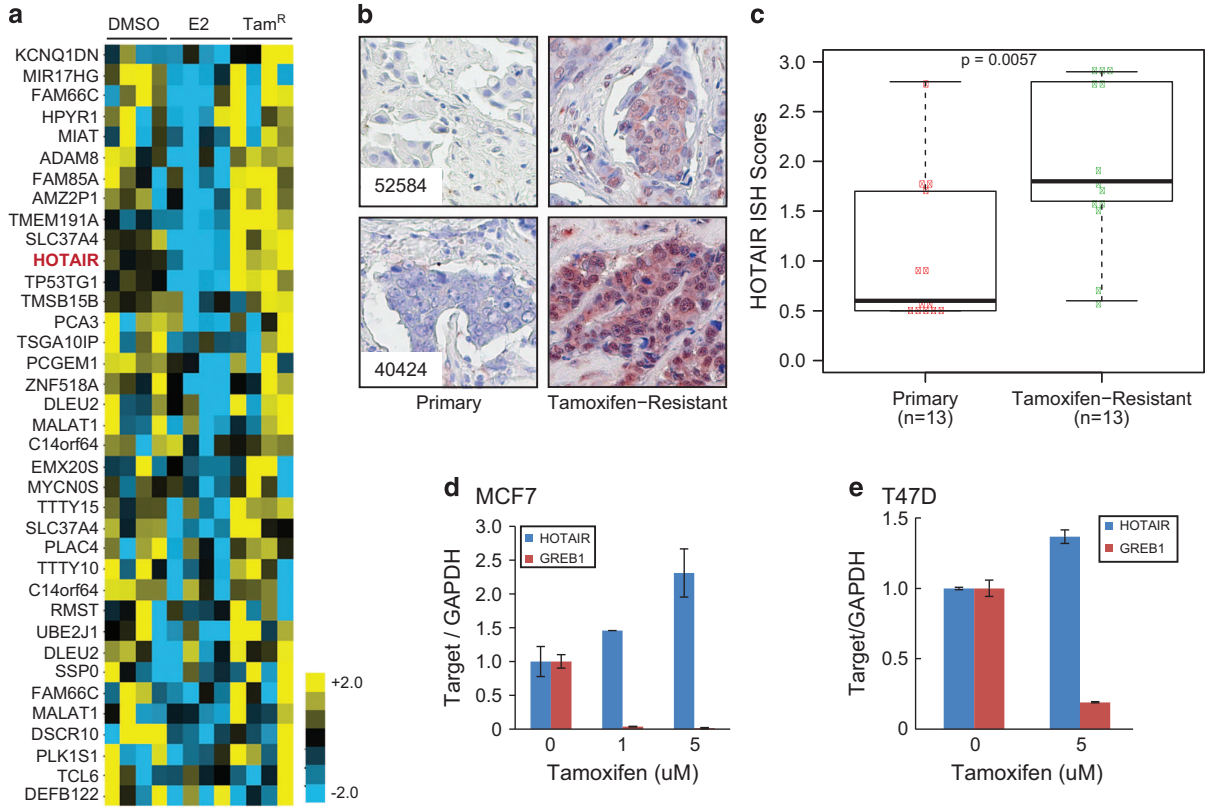



**Figure 1.** *HOTAIR* is upregulated in tamoxifen-resistant (TamR) breast cancer. (**a**) Heat map showing lncRNAs that are repressed by estradiol but upregulated in TamR MCF7 cells. Microarray data were downloaded from GEO with GSE5840 and re-analyzed for lncRNA expression. *HOTAIR* is shown in red. (**b**) *HOTAIR in situ* hybridization (ISH) staining in two representative pairs of primary and TamR breast cancers. (**c**) Boxplot showing *HOTAIR* ISH staining intensity in a set ($n = 13$) of matched primary and TamR breast cancers. (**d, e**) *HOTAIR* expression is increased by tamoxifen treatment. Quantitative reverse transcriptase – PCR analysis of *HOTAIR* and *GREB1* was done in MCF7 (**d**) and T47D (**e**) cells treated with increasing doses of tamoxifen for 7 days. Gene expression was normalized to GAPDH. Data shown are mean ± s.e.m. and are representative of at least two independent experiments.

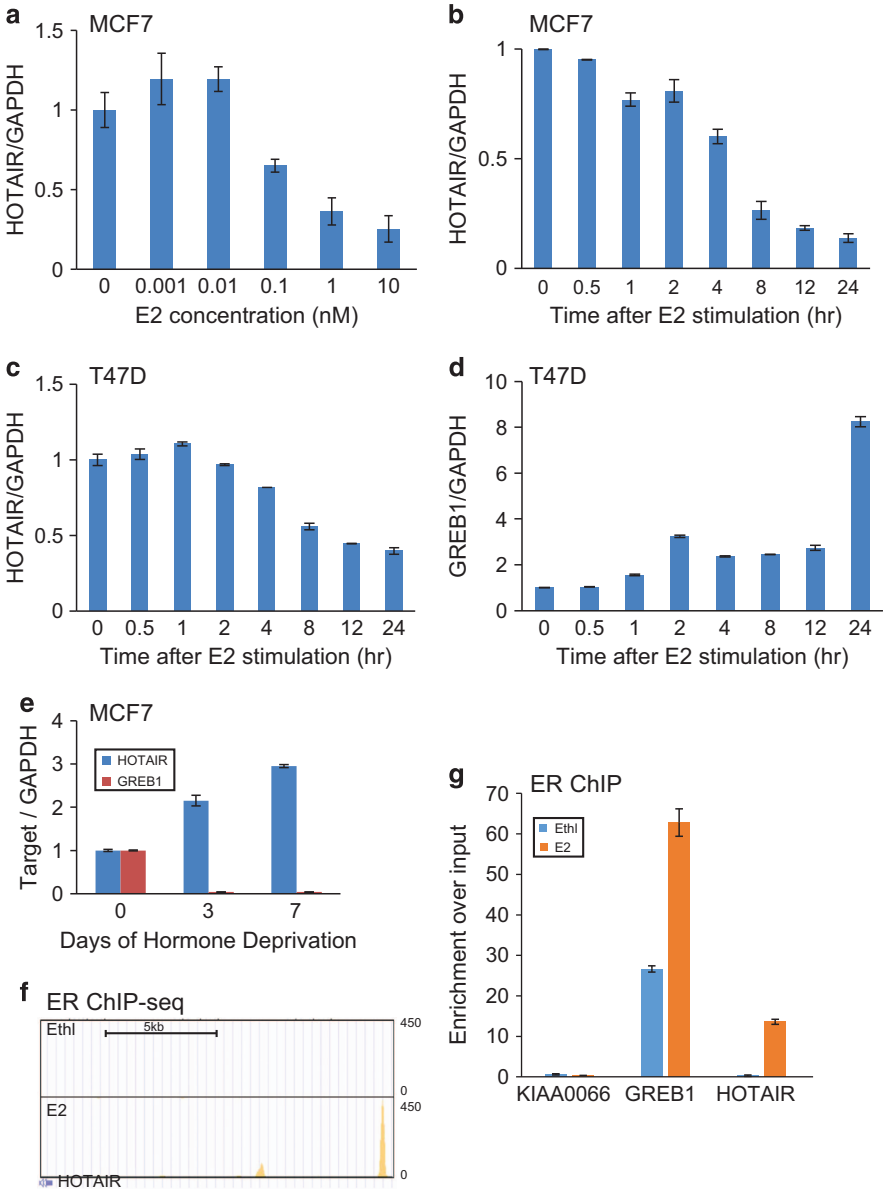
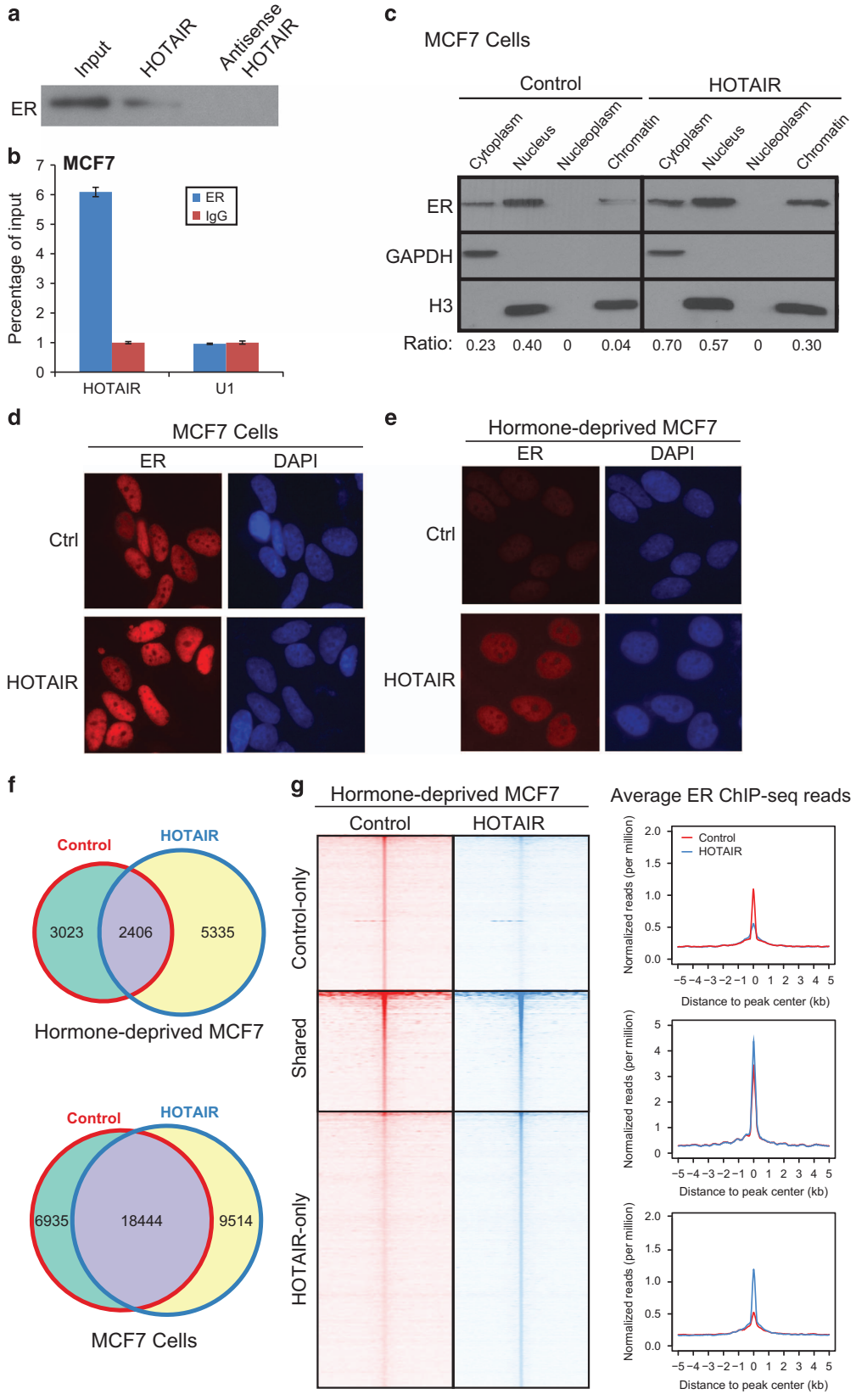



**Figure 2.** The lncRNA *HOTAIR* is directly repressed by estrogen through the estrogen receptor (ER). (**a**) Estrogen inhibits *HOTAIR* expression in a dose-dependent manner. MCF7 cells were hormone starved for 3 days and treated with increasing amounts of estradiol (E2) for 6 h. RNAs were then collected and subjected to quantitative reverse transcriptase – PCR (qRT – PCR) analysis of gene expression. (**b**) Estrogen inhibits *HOTAIR* expression in a time-dependent manner. MCF7 cells were treated with 1 nM E2 and collected at different time points for gene expression analysis by qRT – PCR. (**c, d**) Estrogen inhibits *HOTAIR* expression in T47D cells. T47D cells were hormone deprived for 3 days followed by E2 stimulation for up to 24 h. Cells were then collected at different time-points for qRT – PCR analysis of *HOTAIR* and *GREB1* expression and normalized to GAPDH. (**e**) Estrogen depletion restores *HOTAIR* level. qRT – PCR analysis of *HOTAIR* and *GREB1* in MCF7 cells subjected for hormone deprivation for up to 7 days. (**f**) Genome Browser view of ER binding events at the enhancer of the *HOTAIR* gene. ER chromatin immunoprecipitation sequencing (ChIP-seq) was performed in MCF7 cells stimulated with ethanol (Ethl) or E2[18] and re-analyzed using HOMER (GSE23893). (**g**) ChIP – qPCR showing ER binding to the *HOTAIR* distal enhancer. ER ChIP was performed in hormone-starved MCF7 cells stimulated with ethanol (vehicle) or E2 for 45 min. Enrichment of ER at specific genomic regions including *GREB1* and *HOTAIR* enhancers was evaluated by qPCR. The *KIAA0066* gene was utilized as a negative control as previously described.[34] qPCR, quantitative PCR.

the ER-bound enhancer (P4; Supplementary Figure S2B). Taken together, our data showed that *HOTAIR* is directly repressed by estrogen and is therefore upregulated following hormone deprivation and in TamR breast cancer.

*HOTAIR* directly interacts with ER and enhances ER transcriptional activities

Next, we asked what is the role of elevated HOTAIR in breast cancer ER signaling and tamoxifen resistance. *HOTAIR* has been previously shown to directly interact with chromatin-modifying proteins such as EZH2 and LSD.[15,19,20] As *HOTAIR* is upregulated in TamR breast cancer cells which often have altered ER program,[16] we asked whether *HOTAIR* might regulate ER function. This may shed light on the mechanisms underlying recently reported correlation between *HOTAIR* expression and ER positivity in primary specimens.[9,10] To test this, we first examined whether *HOTAIR* lncRNA could physically interact with the ER protein using RNA pull-down assay. Briefly, we carried out *in vitro* transcription to synthesize biotinylated RNA probes from sense and antisense *HOTAIR* DNA templates, which were then incubated with MCF7




nuclear extracts to allow protein – RNA interactions and precipitated, along with its interacting proteins, with streptavidin beads. Western blot analysis demonstrated that the sense HOTAIR RNA probe, but not the antisense transcript, pulled down the ER protein (Figure 3a). On the other hand, we carried out RNA immunoprecipitation assay and found that the ER antibody

significantly enriched for *HOTAIR*, as opposed to IgG control, whereas the negative control RNA U1 did not exhibit differential enrichment (Figure 3b).

Subsequently, we inquired into the consequences of *HOTAIR*–ER interaction, in order to speculate how ER activities may be affected as a result. By separating MCF7 cell lysates into cytoplasmic, nuclear, nucleoplasm and chromatin-bound fractions, we observed that ER, as expected, localized primarily within the nucleus as opposed to cytoplasm. HOTAIR overexpression substantially increased ER protein levels, suggesting potential roles of *HOTAIR* in enhancing ER transcriptional functions (Figure 3c and Supplementary Figure S3). Moreover, immunofluorescent staining confirmed noticeable increase of nuclear ER following *HOTAIR* overexpression (Figure 3d). Interestingly, this *HOTAIR*-mediated increase in nuclear ER level was also true under hormone-starved condition, in which there is only minimal estrogen present to activate ER translocation into the nucleus, suggesting the roles of *HOTAIR* in enhancing ligand-independent ER function (Figure 3e). To confirm the notion that *HOTAIR* may augment ER genomic targeting, we conducted ER ChIP-seq in MCF7 cells grown in the presence and absence of estrogen. As expected, the total number of ER binding sites was 4.6-fold higher in estrogen-stimulated vs hormone-deprived cells (Figure 3f). Importantly, upon overexpression of *HOTAIR*, global ER binding events were greatly increased under both conditions. Heat map and average intensity analysis of the various groups of ER peaks demonstrated a clear increase in ER ChIP-seq read intensity in both shared and *HOTAIR*-only groups, representing a majority of the ER binding events (Figure 3g and Supplementary Figure S4A). This *HOTAIR*-mediated increase of ER binding events was more prominent in the absence of estrogen, suggesting important functions of *HOTAIR* in regulating ligand-independent ER activities. Concordantly, quantitative PCR analysis of several previously reported ER target genes, such as *GREB1, TFF1, PR*, and *CTSD*, demonstrated that *HOTAIR* overexpression significantly increased ER occupancy at most of these genes (Supplementary Figure S4B). Similarly, the increase in ER binding at target genes was more prominent in hormone-deprived MCF7 cells (Supplementary Figure S4C). Next, we proceeded to investigate to what extent *HOTAIR* impacts ER-mediated transcriptional activities particularly in a hormone-deprived environment.

### *HOTAIR* drives estrogen-independent ER transcriptional program

To identify *HOTAIR*- and estrogen-regulated genes, we conducted microarray profiling of hormone-deprived and estrogen-stimulated MCF7 cells with control or *HOTAIR* overexpression. Data analysis identified 132 and 112 genes that were induced and repressed by *HOTAIR*, respectively. Importantly, hierarchical clustering followed by heat map view revealed that a majority of *HOTAIR*-induced genes are also induced by E2 stimulation, whereas *HOTAIR*-repressed genes tend to become downregulated by estrogen (Figure 4a). Concordantly, gene set enrichment analysis demonstrated that E2-induced genes were significantly upregulated following *HOTAIR* overexpression, even in the absence of estrogen, whereas E2-repressed genes were strongly downregulated by HOTAIR (Figures 4b and c). Gene ontology analysis showed that *HOTAIR*-induced genes were significantly enriched for response to protein stimulus and regulation of cell death and apoptosis, being consistent with the functions of estrogen-mediated ER signaling (Figure 4d; Supplementary Tables S1 and S2). To confirm *HOTAIR* regulation of ER-mediated transcriptional program, we performed quantitative reverse transcriptase – PCR analysis of several known ER-target genes such as *GREB1, TFF1*, and *c-MYC*. Indeed, our data showed that *HOTAIR* overexpression induced ER-target gene expression in the absence of estrogen and further potentiated the effects of E2 (Figures 4e and g). Taken together, we provide evidence for a model by which the lncRNA *HOTAIR* increases ER protein level and enhances its chromatin binding and thus the ER transcriptional program, even in an estrogen-depleted environment. As *HOTAIR* is upregulated in TamR breast cancer, we next asked whether *HOTAIR* contributes to the development of tamoxifen resistance in breast cancer, wherein tamoxifen abolishes estrogen-mediated activation of ER, similar to hormone starvation.

### LncRNA *HOTAIR* promotes tamoxifen-resistant breast cancer progression

To determine the role of *HOTAIR* in breast cancer, we first overexpressed *HOTAIR* in MCF7 cells (Figure 5a). Cell proliferation assay showed that *HOTAIR* overexpression increased MCF7 cell growth (Figure 5b). On the other hand, *HOTAIR* knockdown in T47D cells markedly reduced cell proliferation (Figures 5c and d). To provide direct evidence linking *HOTAIR* to tamoxifen resistance, we generated a TamR MCF7 cell line by continuously culturing the cells in the presence of 5 μM tamoxifen for several months. Consistent with previous *HOTAIR* staining results in TamR breast tumors, *HOTAIR* level showed a remarkable fourfold increase following long-term treatment of tamoxifen (Figure 5e). To determine whether this upregulated *HOTAIR* is critical for the TamR MCF7 cell growth, we performed *HOTAIR* knockdown using two independent short hairpin RNA constructs (Figure 5f). Subsequently, we performed cell proliferation assay to investigate to what extent *HOTAIR* contributes to tamoxifen resistance. As demonstrated in Figure 5g, knockdown of *HOTAIR* significantly decreased TamR MCF7 cell growth, suggesting that tamoxifen resistance may be reverted by targeting or depleting *HOTAIR*. Consistently, clonogenic assays showed that *HOTAIR* knockdown greatly inhibited the colony-formation abilities of the TamR cells, further supporting the role of *HOTAIR* in mediating TamR cell growth (Figure 5h).

## DISCUSSION

With the emergence of studies focusing on the functional attributes of nonprotein-coding RNA transcripts, such as lncRNAs, it has been revealed that these lncRNAs may contribute

**Figure 3.** *HOTAIR* interacts with the estrogen receptor (ER) protein and enhances ER genomic action. (**a**) *HOTAIR* lncRNA interacts with the ER protein. RNA pull-down assay was performed in MCF7 cells using biotin-labeled *HOTAIR* RNA probe transcribed *in vitro*. The antisense *HOTAIR* probe was used as negative control. (**b**) ER protein binds to *HOTAIR* lncRNA. MCF7 cells were subjected to RNA immunoprecipitation assay using an anti-ER antibody or IgG control. Immunoprecipitation-enriched RNA was then analyzed by quantitative reverse transcriptase – PCR. U1 RNA was utilized as a negative control. (**c**) *HOTAIR* overexpression increases ER protein level. MCF7 cell lysates were separated into the cytoplasm, nuclear, nucleoplasm, and chromatin-bound fractions and were detected by western blot analysis. GAPDH and H3 were utilized as loading controls for cytoplasmic and nuclear/chromatin fractions, respectively. Quantification was done by measuring band intensity with ImageJ (NIH, Bethesda, MD, USA) and normalizing to loading control. (**d, e**) Ectopic overexpression of *HOTAIR* increases nuclear ER level. ER immunostaining was performed in control and *HOTAIR*-overexpressing MCF7 cells grown in the presence (**d**) and absence (**e**) of estrogen. (**f**) Overlap of ER-binding sites detected by chromatin immunoprecipitation sequencing (ChIP-seq) in MCF7 cells with control or HOTAIR overexpression in the absence and presence of estrogen. (**g**) Heat map depicting ER ChIP-seq read intensity around (±5 kb) peak centers detected in control or HOTAIR-overexpressing MCF7 cells under hormone-starved condition. Average ER ChIP-seq read intensity around ER binding sites (±5 kb) is shown on the right.

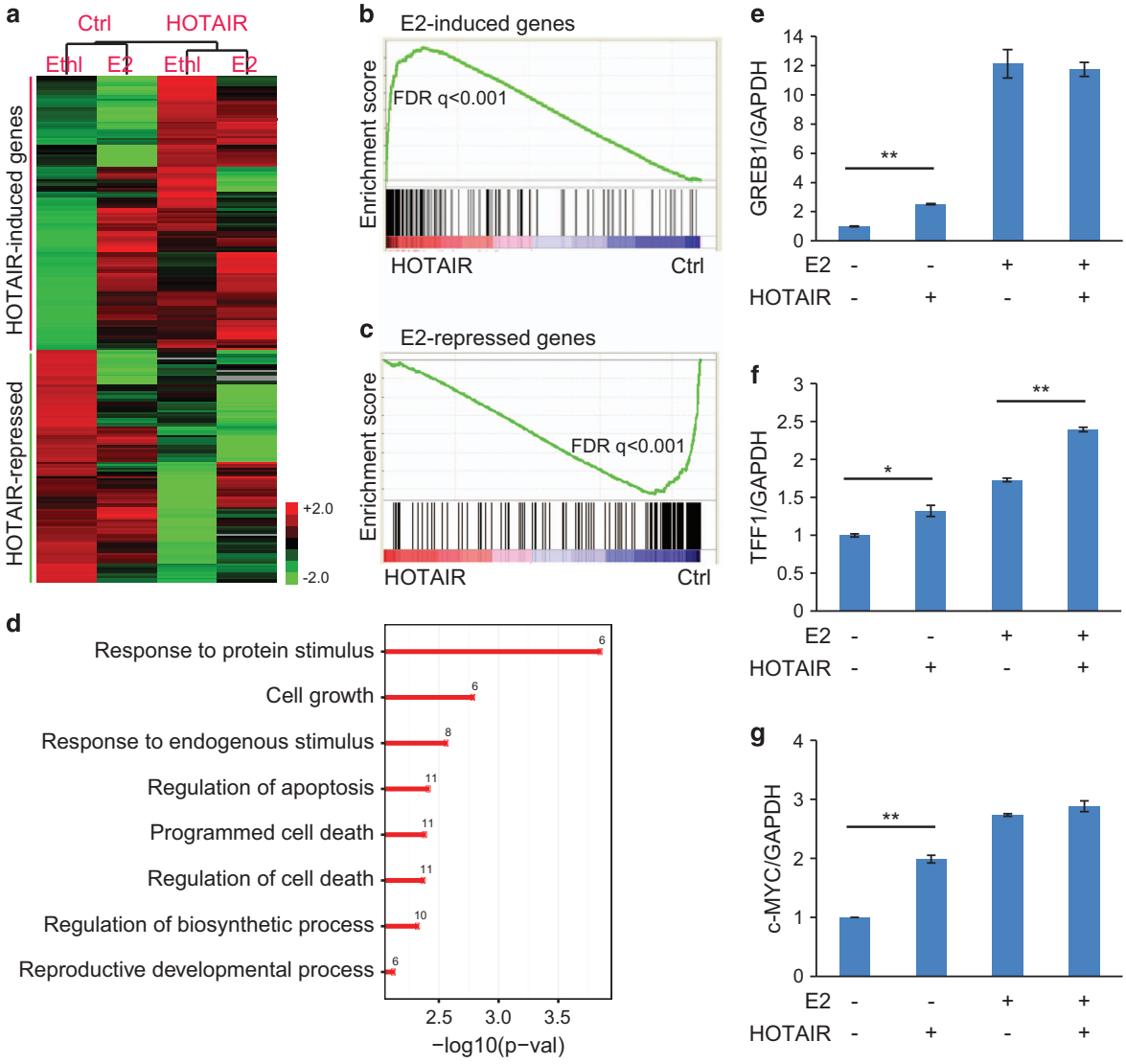



**Figure 4.** *HOTAIR* overexpression enhances estrogen receptor (ER) transcriptional program. (**a**) *HOTAIR*-induced and repressed genes are, respectively, increased and decreased by estrogen. Expression microarray was utilized to profile gene expression in control and *HOTAIR*-overexpressing MCF7 cells that were hormone-deprived for 3 days followed by either ethanol or 1nM estradiol (E2) treatment for 6 h. Expression of genes induced or repressed by *HOTAIR* for at least twofold were clustered and visualized using heat map. (**b, c**) Estrogen-induced genes are significantly enriched for upregulation by HOTAIR (**b**), whereas estrogen-repressed genes are downregulated by *HOTAIR* (**c**). GSEA was carried out to determine the enrichment of E2-induced and -repressed gene sets in the expression data set comparing control and *HOTAIR*-overexpressing MCF7 cells. (**d**) *HOTAIR*-induced genes are enriched for cell growth and response to protein stimulus. Gene ontology (GO) analysis was performed using 132 genes that are increased by *HOTAIR* by at least twofold. Shown here are representative GO terms that are significantly enriched ($P < 0.05$). (**e** – **g**) *HOTAIR* enhances ER-target gene expression. Quantitative reverse transcriptase – PCR analysis of representative ER-induced genes was performed in control and *HOTAIR*-overexpressing MCF7 cells stimulated with ethanol or estrogen. Data were normalized to GAPDH. Error bars: mean ± s.e.m. *$P < 0.05$ and **$P < 0.01$.

significantly to the biological processes involved in physiological as well as pathological conditions. Numerous lncRNAs have been identified as critical players during cancer development; some may be beneficial by acting as tumor or metastasis suppressors (for example, *GAS5*,[21] *MEG3*,[22] *LIFR*[23]), whereas others may be detrimental by promoting oncogenesis (for example, *PCA3* or, as previously named, *DD3*,[24] *PCAT-1*,[25] *SChLAP1*[26]). Previous studies have shown that lncRNAs exhibit great diversity in their functions and mechanisms of action, which include but are not limited to epigenetic transcriptional regulation, association with enhancer and chromosomal looping, and mRNA processing and translation.[27] Several unique properties of lncRNAs make them highly useful in the clinic, with potential utilities including their use as diagnostic biomarkers due to their tissue specificity,[25,28] as well as in lncRNA-based therapies by means of RNA interference.[29] Yet, lncRNAs have just begun to be identified and cataloged; a majority of them remain to be characterized.

Gupta *et al.*[7] reported in 2010 that the lncRNA *HOTAIR* is notably increased in primary breast tumors as well as during metastases. Specifically, by interacting with EZH2 of the polycomb repressive complex 2 complex, which catalyzes trimethylation at histone H3 lysine 27 (H3K27me3) and is upregulated in a variety of aggressive cancers, *HOTAIR* was demonstrated to alter chromatin structure and regulate gene expression, thereby giving rise to an invasive cancer phenotype. In this study, we provide experimental evidence that *HOTAIR* is also critically involved in conferring tamoxifen resistance to MCF7 cells, which represents a major challenge in the clinic

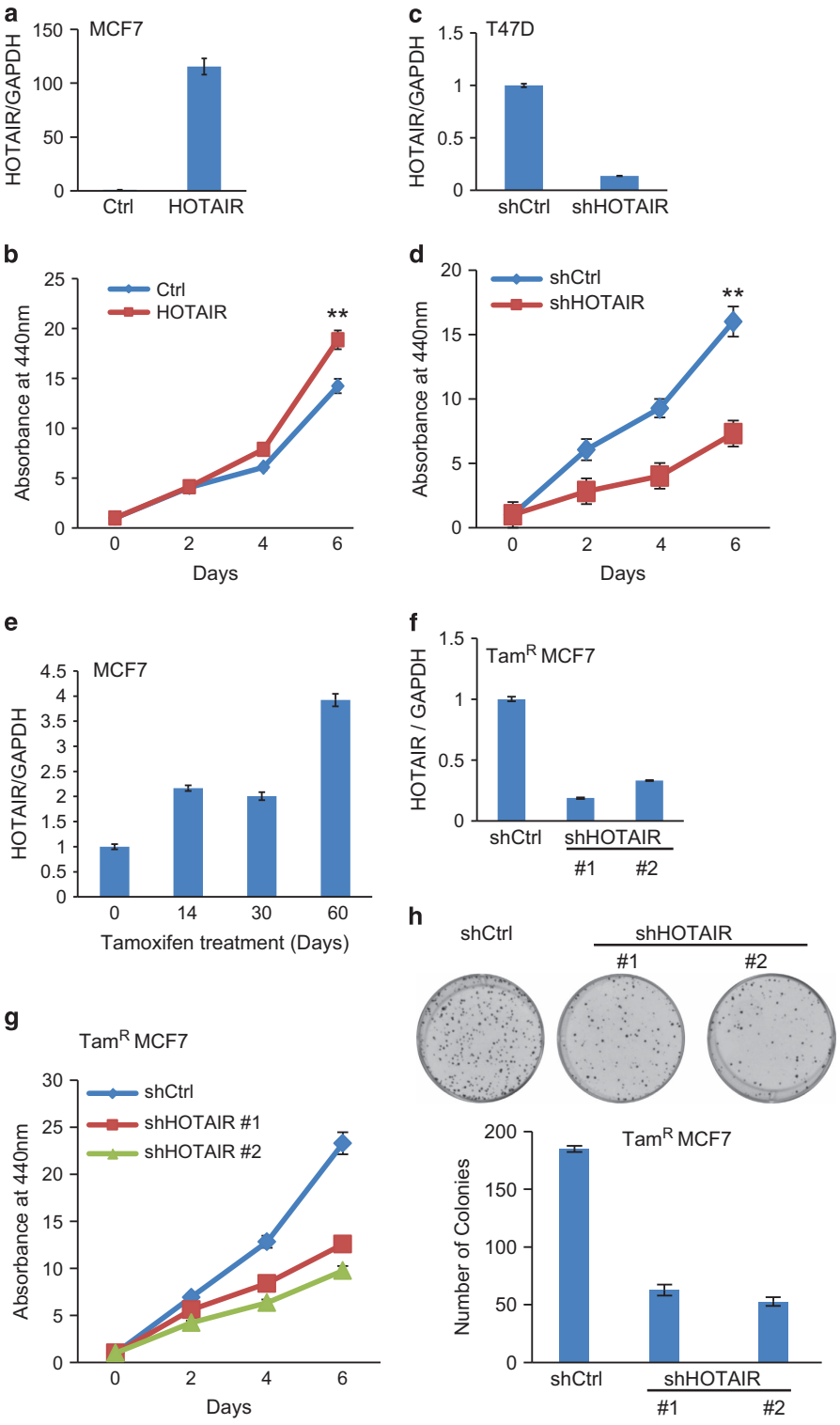



**Figure 5.** *HOTAIR* promotes breast cancer cell growth and tamoxifen resistance. (**a, b**) *HOTAIR* overexpression increases breast cancer cell growth. *HOTAIR* was overexpressed in MCF7 cells through lentiviral transduction, with overexpression confirmed by quantitative reverse transcriptase – PCR (qRT – PCR) (**a**). Cell proliferation was determined by WST-1 cell growth assay (**b**). Data shown are mean ± s.e.m. and are representative of at least two independent experiments, $^{**}P < 0.01$. (**c, d**) HOTAIR knockdown decreases breast cancer cell growth. HOTAIR was depleted in T47D cells through short hairpin RNA (shRNA) lentiviral transduction (**c**) and cell growth was evaluated by WST-1 assay (**d**). Data shown are mean ± s.e.m. and are representative of at least two independent experiments, $^{**}P < 0.01$. (**e**) *HOTAIR* is upregulated in tamoxifen-resistant (TamR) breast cancer cells. MCF7 cells were continuously grown in medium containing 5 μM tamoxifen and periodically harvested for RNA isolation and qRT – PCR analysis. (**f, g**) *HOTAIR* knockdown decreases TamR breast cancer cell growth. *HOTAIR* was depleted in TamR MCF7 cells through lentiviral transduction of two shRNA constructs. Gene expression was determined by qRT – PCR (**f**) and cell growth by WST-1 assay (**g**). (**h**) *HOTAIR* knockdown inhibits colony formation abilities of breast cancer cells. Colony formation assays and quantifications were performed in TamR MCF7 cells stably expressing control or *HOTAIR*-targeting shRNAs.

today. Tamoxifen, belonging to the class of selective ER modulators, is a competitive antagonist of ER that was developed in the 1970s and has been the mainstay treatment for ER-positive breast cancer, which accounts for at least 70% of all breast cancers.[12] Despite its initial success in reducing disease mortality and improving survival, tamoxifen therapy frequently led to the onset of resistance, and recurrence was reported to occur within 15 years in one-third of patients treated with tamoxifen.[12,30] Thus, it has become imperative to understand the mechanisms for acquisition of tamoxifen resistance and to develop targeted therapies to improve treatment for breast cancer.

Our results showed that *HOTAIR* is highly upregulated in the tumors of TamR breast cancer patients compared to their primary tumors before treatment. Moreover, physical interaction between *HOTAIR* and the nuclear hormone receptor ER was detected, which in turn resulted in significant amount of nuclear ER even under estrogen-depleted conditions, thus allowing ER genomic targeting and consequently inducing the ER transcriptional program. Importantly, this phenomenon of *HOTAIR*-mediated activation of ER function in the absence of estrogen indicated a potential route to ligand independence that is manifested in TamR cells. Furthermore, by generating a TamR MCF7 cell model, we showed that *HOTAIR* was consistently upregulated over long periods of drug treatment. In addition, we demonstrated that *HOTAIR* significantly contributes to the growth of these TamR cells. Therefore, in our present study we provide evidence for a novel mechanism that is employed by the lncRNA *HOTAIR* to promote ER activation in the absence of estrogen and drive tamoxifen resistance. Because of this crucial role *HOTAIR* plays in the progression of breast cancer and development of drug resistance, it holds great promise as a useful biomarker and potential therapeutic target.

## MATERIALS AND METHODS

### Patient specimens and cell lines

All breast cancer tissue specimens ($n = 13$) were collected via surgical resection or biopsy from patients diagnosed between January 2006 and February 2014 at the Cancer Center of Guangzhou Medical University. The study protocol was approved by the Ethics Committee of Cancer Center of Guangzhou Medical University. In general, with $n = 10$, for a continuous outcome, there will be $>89\%$ power to reject the null hypothesis of no difference when the difference is 1.5 s.d. or more, using two-sided *t*-test and a type 1 error of 0.05. MCF7 and T47D cell lines were ordered from ATCC.

### Plasmids, reagents, quantitative PCR and western blotting

*HOTAIR* sequence was amplified by PCR and subsequently cloned into the expression vector pCDH-MSCV-mcs-EF1-GFP-T2A-Pu (SBI) at *Eco*R1 and *Not*1 sites using Cold Fusion kit (SBI). The shHOTAIR was cloned into the pLKO lentivirus system. All PCR primers for cloning are listed in Supplementary Table S3 and high-fidelity enzyme Phusion was used for PCR amplification. All PCR products were verified by DNA sequencing. Specific antibodies used in this work include rabbit ER (06-935, Millipore, Billerica, MA, USA), mouse ER (sc-8002, Santa Cruz, Dallas, TX, USA), mouse GAPDH (ab9484, Abcam, Cambridge, MA, USA) and rabbit H3 (ab1791, Abcam). Other reagents include beta-estradiol (E8875, Sigma, St Louis, MO, USA) and 4-hydroxytamoxifen (H6278, Sigma-Aldrich, St Louis, MO, USA). All primers were designed using primer 3 and synthesized by Integrated DNA Technologies (Coralville, IA, USA; Supplementary Table 3). Quantitative PCR was performed using SYBR Green by StepOne Plus in three technical replicates and significance was determined by two-sided *t*-tests. Each experiment was repeated independently at least two times. Western blotting was carried out using standard protocol and repeated at least two times. Band intensity on western blot was quantified with ImageJ and normalized to each respective control to obtain the ratio of ER protein level.

### 3C assay

The digestion map of commonly used restriction enzymes around the enhancer/promoter region of HOTAIR locus (from −103 to +83 kb) and *Bgl*II was selected for digestion, as *Bgl*II sites show a distribution that will enable appropriate primers to be designed to generate 200 – 350 bp PCR products on re-ligation. All primers are designed based on the forward strand immediately upstream of a *Bgl*II restriction site (Supplementary Figure S2B and Supplementary Table S3). 3C experiments were conducted according to the standard 3C protocol as previously described.[31] Briefly, fixed chromatin of hormone-starved or E2-treated MCF7 cells ($1 \times 10^7$) was digested with *Bgl*II overnight and incubated with 50 units of T4 DNA ligase (10799009001, Sigma-Aldrich) overnight in a volume of 7 ml to keep the DNA concentration at 2 – 3 ng/ml to favor intramolecular ligation.

### RNA pull-down assay

RNA pull-down was performed as previously described.[20] Briefly, biotin-labeled RNAs were transcribed from DNA templates with biotin-UTP, NTP mix, and T7 RNA polymerase (Promega, Madison, WI, USA), treated with RNase-free DNase I (Promega) and purified with RNeasy Mini kit (QIAGEN). Nuclei were extracted from MCF7 cells and resuspended in 1 ml RIP buffer (150 mM KCl, 25 mM Tris pH 7.4, 0.5 mM dithiothreitol, 0.5% NP40, 1 mM phenylmethylsulfonyl fluoride, and protease inhibitor (Roche Complete Protease Inhibitor Cocktail Tablets)), and subsequently subjected to mechanical shearing using a dounce homogenizer. For precipitation assays, fragmented nuclear extract and the RNA probe were incubated at room temperature (RT) for 60 min, and 60 µl of Streptavidin agarose beads (Invitrogen, Grand Island, NY, USA) were added to each binding reaction and further incubated at RT for 1 h. After five times of washing with PBS, samples were boiled in SDS buffer and subjected to western blot analysis.

### RNA immunoprecipitation

RIP protocol was derived from published reports.[20] Briefly, cells were treated with 0.3% formaldehyde for 10 min at 37 °C, then added with glycine to a final concentration of 0.125 M and then incubated at RT for 5 min. Cells were then washed twice in cold phosphate buffered saline and pellet was resuspended in 1 ml of RIPA buffer, which was incubated on ice with frequent vortexing for 30 min. Finally, the nuclear lysate was obtained by centrifugation at 13,000 r.p.m. for 10 min. To obtain bead and antibody complex, 20 µl protein beads were mixed with 1 µg antibody and rotated for 4 h at 4 °C. The complex was added to nuclear lysates and incubated overnight at 4 °C and then incubated with RNase-free DNase I (Promega) at 37 °C for 15 min and proteinase K at 45 °C for 45 min. Lastly, RNA was extracted with 1 ml TRIzol (Invitrogen) and analyzed by quantitative PCR.

### Immunofluorescent staining

Cells were fixed with 4% formaldehyde for 15 min at RT and then permeabilized in 0.1% Triton X-100 for 15 min at RT. Cells were then washed by PBS for three times, followed by incubation with 5% normal goat serum for 30 min at RT. Subsequently, cells were incubated with primary antibody, the anti-mouse ER antibody (Santa Cruz), for 1 h at RT. After washing three times with PBS, cells were incubated with secondary antibody, Alexa Fluor 594 goat anti-mouse IgG (Invitrogen), for 1 h at RT. Finally, cells were washed three times with PBS and mounted using Prolong Gold Antifade Reagent (Invitrogen).

### LncRNA *in situ* hybridization

Biotin-labeled antisense HOTAIR RNAprobe /5Biosg/G+C+C+TTGCTCCCTT +G+CCTGCATTTCT+C+T+G was synthesized by EXIQON (Woburn, MA, USA). For paraffin-embedded tissue, after deparaffinization and rehydration, the samples were treated with peroxidase-quenching solution; proteinase K was added to digest tissues before prehybridization and hybridization, which were carried out at 56 °C for 30 min and 4 h, respectively. Then streptavidin–horseradish peroxidase was used to react with the bound biotin-labeled probe. The signal was further amplified using TSA amplification kit (PerkinElmer, Waltham, MA, USA). Finally, the signal was revealed with Ultra Vision One polymer and aminoethyl carbazole chromogen (Thermo Fisher Scientific, Waltham, MA, USA). The stainings were then scored by eye by two pathologists, on a three-tiered scoring system, using the following criteria for the three tiered system:

0 = negative, 1 = equivocal/uninterpretable, 2 = weak positive and 3 = strong positive.

### Gene expression microarray and data analysis

Total RNAs were isolated using TRIzol reagent (Invitrogen). The integrity of the RNA was verified using Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). Microarray profiling was performed using HumanHT-12 v 4.0 Expression BeadChip (Illumina, San Diego, CA, USA). Bead-level data were processed using GenomeStudio (Illumina), and the expression values were quantile-normalized using the limma package in Bioconductor.[32]

Genes having at least twofold changes in *HOTAIR*-overexpressing ells compared with the control cells in the absence of estrogen were defined as *HOTAIR*-regulated gene set. Genes with at least twofold change between ethanol and estrogen-stimulated MCF7 cells were defined as estrogen-regulated genes. Gene ontology terms enrichment was analyzed using DAVID 6.7.[33] Gene set enrichment analysis was performed as previously described.[34]

### ChIP and ChIP-seq

ChIP experiments were carried out as previously described.[35] Antibodies used are Rabbit ER (06-935, Millipore) and Rabbit IgG (sc-2027, Santa Cruz). ChIP-quantitative PCR enrichment of target loci was normalized to input DNA and reported as % input ± s.e.m. ChIP DNA was prepared into libraries according to standard protocols using Bioo Scientific's DNA Sample Kit (cat. no. 514101, Austin, TX, USA). Libraries were sequenced using Illumina Hi-Seq platforms. Sequence reads were aligned to the Human Reference Genome (assembly hg19) using Burrows – Wheeler alignment tool (bwa) version 0.6.1.[36] Microarray and short-read sequencing data have been deposited in the GEO database with the accession number GSE61270.

### Cell proliferation and clonogenic assay

Cell proliferation assay was carried out using the WST-1 kit according to the manufacturer's instruction (Clontech, Mountain View, CA, USA). Briefly, 5000 cells were seeded in a 24-well plate. After adding 50 µl WST-1 reagents per well, cultures were incubated for 2 h and the absorbance at a wavelength of 440 nm was determined using a microplate reader. For clonogenic assay, 500 cells were plated in each well of a six-well plate. When there was visible colony by naked eye, cells were fixed with 4% formaldehyde and were stained with crystal violet (0.25%). Colonies were then counted.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

Bioinformatic analysis was supported by the computational resources and staff contributions provided for the Quest high-performance computing facility at Northwestern University, which is jointly supported by the Office of the Provost, the Office for Research, and Northwestern University Information Technology. This work was supported by the the US Department of Defense W81XWH-13-1-0319 (to JY) and the Research Scholar Award RSG-12-085-01 (to JY) from the American Cancer Society. JK was supported in part by the National Institutes of Health Training Program in Oncogenesis and Developmental Biology (T32 CA080621), and YAY was supported in part by the National Institutes of Health/National Cancer Institute training grant T32 CA009560.

## REFERENCES

1 Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet* 2009; **10**: 155–159.
2 Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell* 2009; **136**: 629–641.
3 Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Mol Cell* 2011; **43**: 904–914.
4 Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* 2009; **23**: 1494–1504.
5 Yoon JH, Abdelmohsen K, Kim J, Yang X, Martindale JL, Tominaga-Yamanaka K *et al.* Scaffold function of long non-coding RNA HOTAIR in protein ubiquitination. *Nat Commun* 2013; **4**: 2939.
6 Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* 2015; **47**: 199–208.
7 Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 2010; **464**: 1071–1076.
8 Gupta S, Iljin K, Sara H, Mpindi JP, Mirtti T, Vainio P *et al.* FZD4 as a mediator of ERG oncogene-induced WNT signaling and epithelial-to-mesenchymal transition in human prostate cancer cells. *Cancer Res* 2010; **70**: 6735–6745.
9 Chisholm KM, Wan Y, Li R, Montgomery KD, Chang HY, West RB. Detection of long non-coding RNA in archival tissue: correlation with polycomb protein expression in primary and metastatic breast carcinoma. *PLoS One* 2012; **7**: e47998.
10 Sorensen KP, Thomassen M, Tan Q, Bak M, Cold S, Burton M *et al.* Long non-coding RNA HOTAIR is an independent prognostic marker of metastasis in estrogen receptor-positive primary breast cancer. *Breast Cancer Res Treat* 2013; **142**: 529–536.
11 Ring A, Dowsett M. Mechanisms of tamoxifen resistance. *Endocr Relat Cancer* 2004; **11**: 643–658.
12 Musgrove EA, Sutherland RL. Biological determinants of endocrine resistance in breast cancer. *Nat Rev Cancer* 2009; **9**: 631–643.
13 Doisneau-Sixou SF, Sergio CM, Carroll JS, Hui R, Musgrove EA, Sutherland RL. Estrogen and antiestrogen regulation of cell cycle progression in breast cancer cells. *Endocr Relat Cancer* 2003; **10**: 179–186.
14 Shang Y, Hu X, DiRenzo J, Lazar MA, Brown M. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* 2000; **103**: 843–852.
15 Shiau AK, Barstad D, Loria PM, Cheng L, Kushner PJ, Agard DA *et al.* The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* 1998; **95**: 927–937.
16 Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 2012; **481**: 389–393.
17 Fan M, Yan PS, Hartman-Frey C, Chen L, Paik H, Oyer SL *et al.* Diverse gene expression and DNA methylation profiles correlate with differential adaptation of breast cancer cells to the antiestrogens tamoxifen and fulvestrant. *Cancer Res* 2006; **66**: 11954–11966.
18 Joseph R, Orlov YL, Huss M, Sun W, Kong SL, Ukil L *et al.* Integrative model of genomic factors for determining binding site selection by estrogen receptor--alpha. *Mol Syst Biol* 2010; **6**: 456.
19 Kaneko S, Li G, Son J, Xu CF, Margueron R, Neubert TA *et al.* Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. *Genes Dev* 2010; **24**: 2615–2620.
20 Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F *et al.* Long non-coding RNA as modular scaffold of histone modification complexes. *Science* 2010; **329**: 689–693.
21 Mourtada-Maarabouni M, Pickard MR, Hedge VL, Farzaneh F, Williams GT. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene* 2009; **28**: 195–208.
22 Zhou Y, Zhong Y, Wang Y, Zhang X, Batista DL, Gejman R *et al.* Activation of p53 by MEG3 non-coding RNA. *J Biol Chem* 2007; **282**: 24731–24742.
23 Chen D, Sun Y, Wei Y, Zhang P, Rezaeian AH, Teruya-Feldstein J *et al.* LIFR is a breast cancer metastasis suppressor upstream of the Hippo-YAP pathway and a prognostic marker. *Nat Med* 2012; **18**: 1511–1517.
24 Bussemakers MJ, van Bokhoven A, Verhaegh GW, Smit FP, Karthaus HF, Schalken JA *et al.* DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res* 1999; **59**: 5975–5979.
25 Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC *et al.* Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol* 2011; **29**: 742–749.
26 Prensner JR, Iyer MK, Sahu A, Asangani IA, Cao Q, Patel L *et al.* The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat Genet* 2013; **45**: 1392–1398.
27 Prensner JR, Chinnaiyan AM. The emergence of lncRNAs in cancer biology. *Cancer Discovery* 2011; **1**: 391–407.
28 Lee GL, Dobi A, Srivastava S. Prostate cancer: diagnostic performance of the PCA3 urine test. *Nat Rev Urol* 2011; **8**: 123–124.
29 Davis ME, Zuckerman JE, Choi CH, Seligson D, Tolcher A, Alabi CA *et al.* Evidence of RNAi in humans from systemically administered siRNA via targeted nanoparticles. *Nature* 2010; **464**: 1067–1070.
30 Clarke R, Leonessa F, Welch JN, Skaar TC. Cellular and molecular pharmacology of antiestrogen action and resistance. *Pharmacol Rev* 2001; **53**: 25–71.

31 Wu L, Runkle C, Jin HJ, Yu J, Li J, Yang X *et al*. CCN3/NOV gene expression in human prostate cancer is directly suppressed by the androgen receptor. *Oncogene* 2014; **33**: 504–513.
32 Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi WX *et al*. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; **43**: e47.
33 Huang, da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; **4**: 44–57.
34 Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005; **102**: 15545–15550.
35 Yu J, Yu J, Mani RS, Cao Q, Brenner CJ, Cao X *et al*. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell* 2010; **17**: 443–454.
36 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* 2009; **25**: 1754–1760.

Supplementary Information accompanies this paper on the Oncogene website (http://www.nature.com/onc)